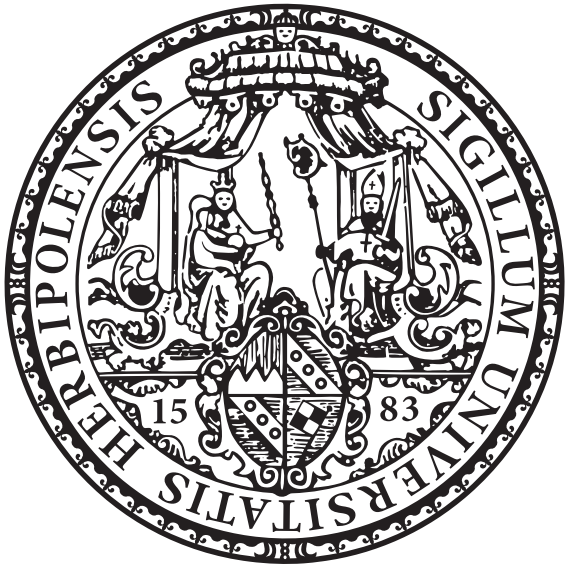


Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität-Würzburg

Theoretical Investigations on the Interactions of Small Compounds with their Molecular Environments

Thomas Christian Schmidt

August 27, 2015

Theoretical Investigations on the Interactions of Small Compounds with their Molecular Environments

Dissertation
zur Erlangung des naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität-Würzburg

vorgelegt von

Thomas Christian Schmidt

aus Schweinfurt

Würzburg 2015

Eingereicht bei der Fakultät für Chemie und Pharmazie am

27. August 2015

Gutachter der schriftlichen Arbeit

1. Gutachter:	Prof. Dr. Bernd Engels
2. Gutachter:	Prof. Dr. Christoph Sotriffer

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer:	Prof. Dr. Bernd Engels
2. Prüfer:	Prof. Dr. Christoph Sotriffer
3. Prüfer:	Prof. Dr. Volker Engel

Datum des öffentlichen Promotionskolloquiums

24. Februar 2016

Doktorurkunde ausgehändigt am

_____

There are naive questions, tedious questions, ill-phrased questions, questions put after inadequate self-criticism. But every question is a cry to understand the world. There is no such thing as a dumb question

- Carl Sagan -

In Gedenken an Carlo und Mona

Contents

1 Introduction

Within this work, two aspects of using theoretical methods for describing the interactions between small molecules and larger biomolecular systems will be presented. One the one hand, this will be the rational design process of covalent inhibitors making use of combining accurate quantum mechanical methods with the capabilities of docking procedures for estimating and optimizing binding affinities of chemical compounds towards their target enzyme. On the other hand, theoretical investigations on the electronic structure of a potential drug molecule and the influences of diverse environments will be shown.

The development of covalent inhibitors is guided mainly by the idea of increasing the residual time of a drug molecule within the active site of its target by maximizing the binding free energy. Without covalency between the inhibitor and its target, the binding free energy is mainly determined by non-covalent interactions such as hydrogen bonds, electrostatic or van der Waals interactions. The magnitude of the resulting interaction potential scales almost linearly with the interaction surface between the inhibitor and its target. Therefore, an optimization of the chemical structure for maximal binding free energies would require the largest interaction surface between the two compounds. However, an analysis of several know inhibitors and their target enzymes revealed, that the binding free energies do only increase up to a system size of about 30 non-hydrogen atom before converging to a constant value of about -60$\frac{kJ}{mol}$, independent of the molecules size.[1] The reason for this limitation can be found the the entropy of the inhibitor, which can move freely while being solvated but receiving significant constraints upon binding to the target molecule. The larger the molecule is, the stronger these constraints affect the binding free energy, resulting in the observed maximum value of -60$\frac{kJ}{mol}$. A further increase of the binding free energies could only be reached by introducing additional exothermicity into the binding process. A possible means for that could be to introduce a chemical reaction, preferably the formation of a covalent bond between the inhibitor and the protein.[2] This bond would anchor the inhibitor while a favorable reaction energy could render the dissociation of the enzyme-inhibitor complex very unlikely or even impossible. The irreversibility however has led concerns against using inhibtors with a covalent mode of action as the high reactivity of some functional groups required for the bond formation also increase the potential of undesired side reactions.[3] Hence the criterion for

the reactivity of a covalent inhibtor to be successful is to be sufficiently large, such that the bond formation with the target protein can happen, and simultaneously as small as possible such as to minimize the reactivity against other host enzymes and such. The theoretical methods used in the design process of such inhibitors are combined quantum mechanical / classical mechanical calculations, which are used to investigate the reaction mechanisms and energies of potential lead compounds in combination with docking procedure, which are used to maximize the binding affinity via and optimization of the chemical structure of the inhibitor. The design process is started from a chemical compound with known covalent inhibition potency and an available crystal structure of the inactivated enzyme-inhibitor complex. QM/MM calculations and docking procedures as then used alternatively in an iterative procedure to study the inhibition reaction and energies and to further optimize the chemical structure until a compound of satisfactory inhibition potency is found. Every time the reactivity against the target is theoretically confirmed, the proposed structures are to be evaluated experimentally.

The second part of this work will be concentrated on interactions between small molecules and their environments. These interactions are routed in their electron densities which, according to the Hohenberg-Kohn theorem,[4] determines all molecular properties. Furthermore, the electron density is accessible by means of X-ray spectroscopy, a field in which great advances have been achieved throughout the last century.[5,6] With the technological advances, ever growing system sizes are accessible. From small organic molecules[7] over biologically active compounds[8,9] the determination of the three dimensional structure of proteins, has evolved to a widely spread procedure with more than 100.000 structures being deposited at the RCSB database[10,11] up to now. Nevertheless, determining the structure of large biochemical systems is still very challenging and, with some rare exceptions,[12,13,14] the resolution remains limited to the position of atoms. Finer details of the electron density cannot be obtained except for small organic and inorganic compounds.[15] Several small, biologically active compounds have been found to arrange comparably in complexes with enzymes and within the crystals of the pure compound concerning the amount and nature of the interactions with the surrounding.[16,17,18,19] These findings and the assumption that the comparable interaction patterns emerge from an optimal arrangement between the compound and its surrounding, has led to the idea of using the crystal structures of the pure compound as estimates to study complexes with their target enzymes. Although the geometries and orientations of small organic molecules have been shown to correspond well between structures taken from the Cambridge structural data bank and the RCSB protein data bank,[20] only few investigations comparing the corresponding electron densities could be found. The selected system for these investigations is again a model for a covalent inhibitor. The chemical envi-

ronments that will be studied are the complex of the molecule with its target enzyme taken from theoretical studies, the crystal structure of the pure compound which was experimentally determined[21] as well as further theoretical model treating the isolated molecule in vacuum or simulating aqueous solution. The electron densities for the inhibitor will be computed by using density functional theory either directly, for the isolated molecule, or within the framework of QM/MM calculations, which allows to take the atomic structure of the environment for the crystal and protein-based systems into account.

2 Theoretical Methods

The theoretical methods used within this work comprise well established theories for the description of molecular systems as well as methods for the analysis of special properties of them. The structure of the investigated systems is thus determined using quantum mechanical methods, classical mechanics or a combination of them. The same methods are furthermore used to determine chemical reactivities, relative energies of the participating compounds as well as the electron distribution within the systems. The obtained electron densities for the model systems were further analysed using the quantum theory of atoms in molecules.[22,23,24]

2.1 Classical mechanics

The most simple description for molecular systems is the idea of the ball-and-spring model, in which the atoms are treated solely by classical mechanics. Bonded interactions between the individual atoms are described by harmonic potentials, for bond distances and bond angles, or periodic potentials, in the case of torsions around single bonds. Non-bonded interactions are covered by classical electrostatics using the atomic charges and van der Waals interactions, usually described via a Lennard-Jones potential.[25] The major difference between the available force fields is found in the corresponding parameters, which are usually derived and optimized according to experimental data. Calculations within this work were carried out using the AMBER[26] and Charmm[27] force field, both of which have been optimized for proteins in aqueous solution. Preparation and analysis of these calculations was performed within VMD,[28] wereas the actual calculations have been carried out using NAMD.[29]

2.2 Electronic Structure Calculations

If not only the molecular geometries but the details of the electronic structure of the model system is of interest, electronic structure methods are mandatory. These methods can be divided into wavefunction based theories such as the Hartree Fock theory and derivatives thereof as well as density functional theory methods. A detailed description of these methods will not

be given here and the interested reader is referred to the corresponding textbooks.[30,31,32,33,34] Among the multitude of methods that are available, only a few have been used within this work. Due to its simplicity and well established implementation in most quantum mechanical program packages, the Hartree Fock method was used for benchmark calculations to validate proper input files especially for the later AIM analysis. Besides the benchmark calculations, however, neither Hartree Fock nor post Hartree Fock calculations were used within this work. The majority of calculations on the electronic structure of molecules have been conducted using density functional theory using either the B-LYP[35,36] or the B3-LYP[35,36,37,38] functional. Despite the development of newer functionals[39] with improved performance for specific classes of molecules, those employed within this work have been found to be sufficiently accurate for the description of the investigated systems.[40,41] A further advantage especially for the B-LYP functional is the efficient implementation within Turbomole,[42] which in connection with the RI approximation[43] can drastically reduce the computational demands. Whereas geometry optimizations have been carried out using B-LYP, the electronic structure as well as relative energies have always been determined by using B3-LYP due to its increased accuracy.[40,41] For all calculations involving electronic structure calculations, a triple-ζ basis set with polarization functions (TZVP) was used.[44] Benchmark calculations have shown that the riple-ζ quality provides the best ratio between computational demands and obtained energies while the additional polarization functions have been found to be needed in the description of charged species and for calculating molecular dipole moments.

Although being feasible for calculations on individual structures, computations on density functional level have been found to be too demanding for scanning the conformational space of some of the model compounds. Therefore, a much less demanding semiempirical method, PM7[45] was used, which is implemented in the MOPAC program package. Calculations using MOPAC were conducted in connection with the CAST[46] program, which internally interfaces MOPAC.

For electron structure calculations of molecules that require an explicit treatment of the environment, for example reactions within the active site of an enzyme, combined quantum mechanical / classical mechanical calculations (QM/MM) have been used.[47] These calculations divide the model system into a small part with its electronic structure explicitly being computed using density functional theory while the remaining part of the system is covered by classical mechanics in form of a force field. The interactions between the two subsystems are covered by the electrostatic embedding scheme.[48] Covalent bonds along the boundary between the QM- and MM-subsystem are treated by using link-atoms.[49,50] QM/MM calculations within this work have been carried out using the Chemshell package[51] which interfaces Turbomole for the quantum mechanical calculations. The force field contribu-

tions are computed internally using the DL_POLY code.[52]

2.3 Quantum Theory of Atoms in Moelcules

Molecules are constructed from atom and functional groups and, as every undergraduate chemistry student learns, functional groups have characteristic properties and are transferable between different molecules. Whereas this simple idea has proven useful in experimental chemistry and shown its value as basis for group additivity schemes, its origin had long time not been understood. It took until the 80ies when Bader introduced the Quantum Theory of Atoms in Molecules (QTAIM or AIM),[22,23] which provided a physically based explanation grounded in the molecular charge distribution, the electron density $\rho(r)$. The shape of this three dimensional function determines the topology of the system as well as a unique partitioning of space into regions corresponding to functional groups or individual atoms. Whereas the interested reader is referred to the literature[24,53] for further information, some of the main concepts are briefly given below.

The central quantity of the AIM theory is the three dimensional electron density distribution which, depending on the system of interest, may exhibit up to four types of stationary points, minima, saddle points of first and second order and maxima. At each of these points, the gradient in $\rho(r)$, $\nabla\rho(r) = \frac{\partial d}{\partial dx} + \frac{\partial d}{\partial dy} + \frac{\partial d}{\partial dz}$ vanishes. To discriminate between the four possible types one has to evaluate the hessian matrix, a $3x3$ matrix of all nine second derivatives of $\rho(r)$. As the hessian matrix is necessarily real and symmetric, it can readily be diagonalized resulting in its eigenvalues λ_{1-3}, representing the curvature along the principle axis of a critical point, represented by its eigenvectors. The number of non-zero eigenvalues, the rank (ω), and the algebraic sum of their signs, the signature (σ), can be used to classify the critical point according to Table 2.1.

ω	σ	type	name and abbreviation
3	+3	minimum	cage critical point, CCP
3	+1	1^{st} order saddle point	ring critical point, RCP
3	-1	2^{nd} order saddle point	bond critical point, BCP
3	-3	maximum	nuclear critical point, NCP

Table 2.1: Mathematical type of citical points determined by the combination of rank (ω) and signature (σ) and chemical interpretation thereof.

Whereas the electron density in molecular systems exhibits cusps of infinitely large values at the position for the nuclei and is not differentiable,

these points behave topologically as if they were true maxima and are usually also recognized as such by computer programs.[53] Also present in any multi-atomic system are bond critical points, which are found on the path of maximal electron density connecting two nuclei. BCPs have been used extensively in discussing the nature of chemical interactions[54,55,56,57,58] as well as in deciding about the existence of absence of a chemical bond.[59,60,61,62] Despite the great interest in BCPs, ring and cage critical points have only received minor attention and also the analysis carried out within this work will focus on BCPs and quantities computed thereat.

The quantities investigated at BCPs encompass the electron density ρ_b, the Laplacian of the electron density $\nabla^2\rho_b$, being computed as trace of the diagonalized hessian matrix, and the ellipticity of the bond, $\epsilon = \frac{\lambda_1}{\lambda_2} - 1$. The absolute value of the electron density is used as a measure for the strength of the interaction between the participating atoms, fragments or molecules. The Laplacian $\nabla^2\rho$ provides a measure of the enrichment or depletion of electron density and, furthermore, a characterization of the bond with respect to covalent or ionic interactions. The last quantity, the bond ellipticity ϵ reflects the symmetry of the electron density distribution along the bond path, asuming values around 0 indicating rotational symmety for single bonds, and reaching up to 0.45 for the prototypical double bond in ethane.[53]

2.4 Description of Molecular surfaces

For the investigations on interactions between separate molecules, it is necessary to define an interface between the participating species. Especially in the study of internal and external influences on a molecule, space has to be partitioned into two regions "inside" and "outside" the molecule, thus requiring a molecular surface as boundary between them. Though conceptually straight forward, there is no such thing as a molecular surface for real systems, as the electron density continually decreases with larger distances from the nuclei, yet it never reaches zero. However, the idea of a molecular surface is very helpful for many aspects in theoretical chemistry, thus there are several diverse methods reported in the literature how to obtain such a model. The earliest approaches to construct molecular surfaces are based on the idea of van der Waals radii of individual atom, constituting the smallest distance that atoms can approach each other without chemically reacting. Measurements of interatomic distances in numerous crystal structures have been used to derive a more or less consistent table of van der Waals radii for the elements.[63,64,65] Lee and Richards[66] used these van der Waals radii to estimate surfaces of biomolecules, namely the solvent accessible and the solvent excluded surface. Both surfaces can be derived using the "rolling ball" algorithm reported by Shrake and Rupley.[67] In this algorithm, a sol-

vent molecule, approximated by a sphere of defined radius, is moved in such a ways, that its perimeter is always in contact with at least on of the van der Waals surfaces of the moleculs atom, but it is never allowed to overlap with one of them. Tracing the center of this sphere, the resulting surface is termed solvent accessible surface which is used widely in computations on the free energy of biomolecules and the prediction of their tertiary structure. Instead of using the center of the sphere, mapping the points of contact between the probe and the van der Waals radii of the atoms results in the solvent excluded sphere, which gives a much better approximation of the boundary between "inside" and "outside" of a molecule than the solvent accessible surface. Whereas the surfaces are widely used for investigations on biomolecules, they are not able to represent the finer details in the electronic structure. Hence, methods were developed using the electron distribution of a molecule as criterion to define the molecules surface. One of these methods is the theory of atoms in molecules reported by Bader,[22,23,24] which uses the zero flux surfaces between adjacent atoms as definition for the atomic basins and thus also for the molecular surface. Whereas this theory has experience broad application in the analysis of crystal structures, it can not provide molecular surfaces for isolated compounds. As the electron density decays continuously at distance from the molecule, there gradient never vanishes and thus no zero flux surfaces encompassing the system can be found. A further problem of the Bader surfaces is the shape of the molecular surfaces, which may lead to technical difficulties while computing properties by integration of the corresponding densities. Another electron density based method to derive molecular surfaces was reported by Hirshfeld[68] and it has found extensive application in the field of crystallography and the analysis of crystal structures.[69,70,71] To derive the shape of the Hirshfeld surface, the electron density is decomposed into contributions arising from each individual atom $\rho_a(r)$. To estimate the contributions, spherically averaged ground-state densities are used. A weighting function can then be used to estimate the amount to which the molecule or its surrounding contribute to the overall electron density at each point in space.

$$w(r) = \frac{\sum_{mol} \rho_a(r)}{\sum_{total} \rho_a(r)} \tag{2.1}$$

Wherever the value of $w(r)$ exceeds 0.5, the electron density is dominated by contributions arising from inside the molecule of interest, thus the surface of exactly $w(r) = 0.5$ is used to define the Hirshfeld surface of a molecule. In contrast to the Bader surfaces, Hirshfeld surfaces have a much smoother shape. This however causes the Hirshfeld surfaces to not being space filling. Whenever used to partition space not only into a molecule and its surrounding but into multiple molecules, regions with significant contributions of more

than two molecules will not be assigned to any of them. Additionally, Hirshfeld surfaces can also not be derived for isolated systems, as the weighting function in this case will always be 1.0 as the only source of electron density is the molecule itself.

Much simpler than the determination of Bader or Hirshfeld surfaces is to use the amplitude of the electron density as decisive criterion. With the density being highest at the nuclei and rapidly decreasing with larger distance, large values can be assigned to the "inside" an small ones to the "outside" of a molecule. This approach however requires the definition for a threshold value, for which no unambiguous choice can be made. Choosing a value of 0.005 atomic units in the electron density is a common practice to generate the corresponding isodensity surface which is then used as a model for the molecular surface. Surfaces of 0.005 a.u. have shown to agree rather well with the experimentally determined van der Waals radii as well as with Bader or Hirshfeld surfaces derived from crystal structures. However, isodensity surfaces can also be computed for isolated molecules making this type of surface applicable for the investigations on the environmental influences on the electron density described later.

3 Design of Covalent Inhibitors

The aim of rational design of new drug molecules is to develop chemical compounds which show a large binding affinity towards their target protein, preferably also exhibiting a strong selectivity for it, such that side effects due to the reaction with further species is minimized. The usual procedure for maximizing the binding affinity is to search for a lead structure, which can then be modified to perfectly address the target enzyme. From a theoretical point of view, docking procedures provide a helpful means for this process, as they are designed to reliably estimate binding free energies between model compounds and their targets. Nevertheless, docking procedures alone concentrate mainly on non-covalent interactions, such as hydrogen bonds, electrostatic and van der Waals interactions. Whereas these interactions can well be used to describe the binding affinities, they suffer from two major flaws. The estimated binding energies are calculated for a very specific configuration of the target protein as well as the model compound. Nevertheless, due to the high rate of mutations in organisms such as bacteria or viruses, changes in the sequence of their proteins are quite frequent. A well studied example if the HIV-1 protease, which is known to retain its catalytic activity, even if up to 20% of the amino acids are exchanged.[72,73,74,75,76] Nevertheless, even small mutations affecting a single amino acid only might already lead to resistance against inhibitor compounds, which have been highly active against the wild type enzyme. One possibility to solve this issue is to address not only the binding pockets of the target enzyme, but to find a reactive species, capable of forming a covalent bond with one of the catalytically active amino acids. By attaching the new compound to the catalytic center, the protein is no longer functional. The only way to develop resistance against this kind of inhibitor would be to mutate the active site residues, in which case the catalytic activity is also lost. The downside of such compounds is, that they require a relatively large reactivity, sufficient to overcome the activation barrier that needs to be overcome during the bond formation process. A high reactivity though implies a higher probability of possible side reactions that might lead to severe side effects if administered as a drug molecule which in turn has led to concerns about the development of covalent inhibitors. There are, however, some widely used examples of drug molecules, that act by covalent binding, for example esomeprazole (Nexium; AstraZeneca) and clopidogrel (Plavix; Sanofi-Aventis/Bristol-Myers Squibb)[3]. Thus, by careful selection of the reactive group, it should be possible to design inhibitors

that posses both, a high binding affinity through non-covalent interactions and the ability to retain their inhibition potency despite mutations of the target protein.

Another advantage of covalent binding can be found in the energetics of the reaction. The binding energy of strictly non-covalently inhibitors is based solely on non-covalent interactions between the inhibitor and the target enzyme. Nevertheless, similar interactions can also be found between the unliganded enzyme and the surrounding solvent or between the solvent and the inhibitor. Thus the binding free energy is not determined by the interactions in the protein-inhibitor complex alone, but by its difference to the solvated but separated species. Strong interactions, such as hydrogen bonds, can also be formed with surrounding water molecules. Similarly charged groups receive a significant stabilization by surrounding water molecules. With a given binding pattern within the protein-inhibitor complex, one would thus have to reduce the solubility in the surrounding solvent such to increase the difference between the solvated and complexed species. Nevertheless, such attempts are estimated to yield no more than about 60 $\frac{kJ}{mol}$.[1,2] Alternatively, instead of making the solvated species more unfavorable, one could also aim to introduce stabilizing interactions which can not be formed in the solvent alone. One such possibility is the above mentioned formation of a covalent bond between the inhibitor and the target enzyme. Covalent interactions have been shown to stabilize hydrogen bond system in crystals about two- to threefold when compared to "usual" hydrogen bonds[77] and several promising, covalently binding compounds have shown to increase binding affinity by at least three orders of magnitude.[2]

To exploit the advantages of covalent binding during the rational design of new compounds, docking procedures are of limited value, as covalent binding is usually only implemented as the definition of an unbound and a bound state.[78,79] To adequately account for the bond formation process, the docking procedures have to be augmented by the use of quantum mechanical methods, which due to their detailed description of electronic effects, are capable to model bond formation processes. Therefore, in the present chapter I will describe the protocol used to combine the benefits of docking procedures and quantum mechanical methods for the rational design of covalent drugs, illustrated for the example of an aspartic acid protease, the HIV-1 Protease.

3.1 The Target Protein

The model system I used during the development of the new protocol for the rational design of covalent inhibitors is the HIV-1 Protease, as this enzyme exhibits several properties that facilitate its use during the current study. In the first place, the HIV-1 Protease plays an essential part in the repro-

duction cycle of the Acquired Human Immunodefficiency Syndrom (AIDS), which classifies the protein as possible point of attack for new drugs. Several experimental studies have shown, that the active form of the enzyme is required for new virions to be infective.[80,81] Further investigations have shown that the catalytic activity results from two aspartic acid moieties which, upon dimerization of two identical protein chains, together form the catalytic dyad.[82,83] The experimental studies carried out on the HIV-1 Protease were son extended to investigations on possible inhibitors including structure determinations by means of NMR and Xray crystallography.[84,85,86,87] These studies mark the beginning of the structural survey on HIV-1 Protease in its wildtype form as well as diverse synthetically engineered mutants, mostly containing inhibitor model compounds. A selection of almost 50 structures is depicted in Figure 3.1 with the corresponding pdb ID codes and authors provided in Table 3.1. The huge number of available crystal structures together with the numerous publications both on experimental[88,89,90,91] and theoretical[92,93,94,95] results form a solid basis for the ongoing investigations and reference data to compare the results to. As a last point to mention are the successful design of several non-covalent inhibitors, some of which are already approved to be administered by the FDA.[96,97,98,99,100,101]

3.1.1 Structure of the HIV-1 Protease

The structure of the HIV-1 Protease can be described as a c_2-symmetric dimer of two identical, 99 amino acid residues long protein chains. The active site of the protein is composed of each ASP25 moiety of the two protein strands, classifying the HIV-1 Protease as aspartic acid protease. The active site is located in two alpha-helical strands right above the dimerization interface, being composed of a beta sheet structure comprising the N- and C-termini of both monomers. Around the active site, there are four distinct binding pockets (S1-S4), while the top is usually covered by the two flaps, two flexible protein loops with the Ile50/Ile50' moieties being located directly above the active site. As depicted in Figure 3.1, the three dimensional structure of the HIV-1 Protease is extremely insensitive to mutations and the overlay of 44 crystal structures does hardly exhibit any deviations. The only flexibility is found in the two flaps which adopt to the steric demands of complexed molecules. They have been shown to exhibit even larger flexibility in solution, where they can open up comparably to a two-winged door, allowing easy access to the active site.[102,103,104,105,106,107] A notable feature of many of the crystals structures[98] and theoretical investigations[108] is the presence of a "flap water" molecule, a molecule of water, tightly bound to the aminofunctions of ILE50/ILE50' via hydrogen bonds and, where possible, to the side of the complexed molecule opposite to the catalytic dyad. Though usually tightly bound, this molecule of water can readily be replaced

by the inhibitor, if the later provides a suitable functional group to accept two hydrogen bonds from the ILE50/ILE50' residues.[109] The later effect was also investigated due to the possible favorable entropic effect caused by the release of the flap water molecule.[110,111]

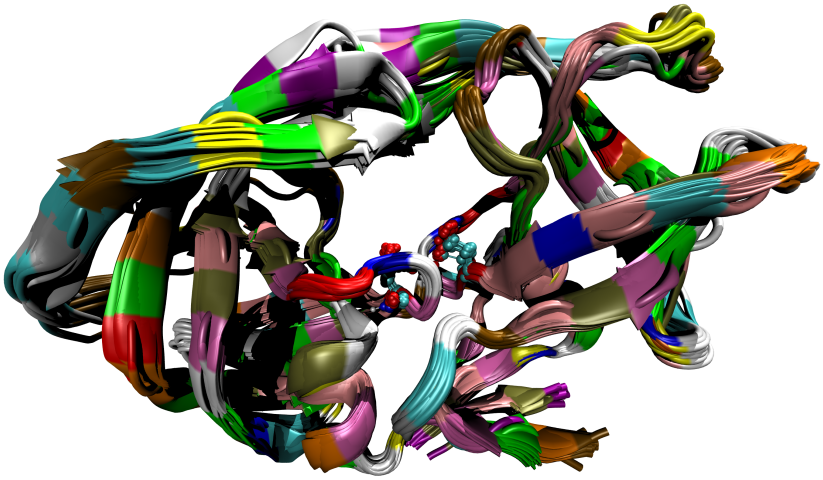


Figure 3.1: Overlay of X-ray crystal structures taken from the RCSB database. xx structures represent the HIV-1 protease in complex with a variety of inhibitors as well as in its uncomplexed form. For comparison, also the crystal structure of SIV protease in complex with EPNP is given to show its structural similarity to the HIV-1 protease. A list of the crystal structures is given in Table 3.1.

3.1.2 Protonation States

For the investigations on the reactivity of HIV-1 Protease, the protonation states of titratable groups play an important role. Nevertheless, there are not many such side chains present. Considering the pH range in which the enzyme usually acts (4-6), only the degree of protonation of the His69/His69' and the one of the catalytic dyad are unclear. Due to the position of the His69/His69' residues on the outer perimeter of the protein, their protonation state is expected to have an only negligible influence on the catalytic activity within the active site. The one of the ASP25/ASP25' residues however has a significant impact.

Considering the two side chains of the catalytic dyad, there are three possible protonation states. In highly acidic environment, both amino acid residues could reside in form of the corresponding carboxylic acids, probably being stabilized by formation of two strong hydrogen bonds between

author	structures used in Figure 3.1	additional structures
Rose[112]	1yth	1ytg 1yti 1ytj
Rose[113]	2sam	-
Ekegreen[114]	2cem	2cej 2cen
Klei[115]	2fxe	2fxd
TIE[116]	2ien	2idw 2ieo
Chellappan[117]	2psu 2psv	-
Rutenber[90]	1aid 2aid	-
Rutenber[118]	3aid	-
Andersson[119]	1ebz	1d4i 1d4j 1ebw 1eby 1ebz 1ec1 ec2 1ec3
Skalova[120]	1lzq	-
Thais[121]	2upj	1upj 3upj 4upj
Wu[122]	2uxz	2uy0
Mahalingam[123]	2wkz	2w10
Tyndall[124]	3bxr 3bxs	-
King[125]	3ekp 3ekq 3ekt 3ekv 3ekw 3ekx 3eky	-
	3el0 3el1 3el4 3el5 3el9	
Nalam[126]	3gi4 3gi5 3gi6	-
Gosh[127]	2i7e	3i6o
Ishima[128]	3jvw	3jvy 3jw2
Olajuyigbe[129]	3k4v	-
Kawasaki[130]	3kdb	3kdc 3kd0
Shen[131]	3nu3 3nu4 3nu5 3nu6 3nu9 3nuj 3nuo	-
Bihani[132]	3kt2 3kt5	-
Liu[133]	2g69	-
Spinelli[134]	1hhp	-
Duskova [DOI 10.2210/pdb1zpk/pdb]	1zpk	-

Table 3.1: Author list and pdb reference codes of the crystal structures used in Figure 3.1. pdb codes given as additional structures denote crystal structures of the same group, but without noticeable structural differences to the one given in the first column.

these groups. Nevertheless, given the pH-range for optimal activity and an estimated low (3.3±0.1) and high (6.8±0.1) pK_a value for the catalytic dyad,[135,136] a protonation of both residues is unlikely.

With only a single proton, either of the ASP25/ASP25' moieties could be protonated. Yet, due to the inherent c_2-symmetry of the enzyme, both cases would lead to equivalent structures, which will thus be considered being identical within this work. The monoprotonated form agrees best with the estimated pK_a values. In connection with experimental studies, this protonation state also correlates well with the formation of a cyclic ring structure between the two carboxylic acid functions forming a low barrier hydrogen bond[137,138,139,140] and a molecule of water bridging the remaining carbonyl oxygen atoms (see Figure 3.2).[141,142]

The third possibility would be a completely deprotonated catalytic dyad with both aspartates being present as carboxylates. Although, with respect to the pK_a values, more probable than a completely protonated form, the deprotonated state is expected to be disfavored due to the interactions of the two negative charges repelling each other, supporting the assumption of a monoprotonated catalytic dyad in the free enzyme.

Whereas the protonation state of the free enzyme can rather unambiguously be determined, a common one for the protein-inhibitor complexes is not. With the broad variety of crystal structures of complexes with HIV-1 Protease, the degree of protonation for the catalytic residues ranges from completely deprotonated[73] over the monoprotonated form of the free protein[143] up to completely protonated. The finding of these atypical protonation states[144] lead to development of further molecules, tailored to address these such as piperidine[100] and pyrrolidine[145,146,147] based inhibitors.

3.1.3 Catalytic Cycle and Inhibition Mechanism

The catalytic cleavage of peptide bond by HIV-1 Protease can be decomposed into the two step reaction formula given in equation 3.1. In the first step of the reaction, the separate species enzyme, denoted E, and substrate, denoted S, have to form a preliminary complex labeled $E \cdots I$. The association constant K_i can be used to quantify the binding affinity between the two species. Once formed, the $E \cdots I$ complex marks the beginning of the actual catalytic reaction leading to the degraded products P. This second step of the reaction is drawn as reversible reaction. Nevertheless, in the real system, the rate constant for the bond cleavage, k_{forward}, is so much larger than that of the reverse reaction, k_{back}, such that the bond cleavage can be assumed to be irreversible.

$$E + S \overset{K_i}{\rightleftharpoons} E \cdots S \underset{k_{\text{back}}}{\overset{k_{\text{forward}}}{\rightleftharpoons}} E + P \tag{3.1}$$

While the catalytic cycle involves both steps, binding of the substrate and cleavage of the bond, only the second step comprises chemical reactions. Hence the term "mechanism" is often ambiguously used both for the overall process as well as for the second step of the reaction only. The chemical, second step of the catalytic reaction is based on the prevalent monoprotonated state of the catalytic dyad. Early assumptions of the mechanism involve a direct nucleophilic attack of the deprotonated ASP25 while the second ASP25' acts as general acid/base catalyst [148] merely due to its spatial proximity. [149,150] Later experimental results however indicated, that a direct bond formation between the nucleophilic ASP25 and the substrate might not be possible due to a molecule of water, caught between the ASP25/ASP25' moieties and stabilizing this protonation state as described above. Hence another mechanism was proposed, in which the nucleophilicity of the deprotonated ASP25 residue is transferred onto this molecule of water, which in turn can perform the nucleophilic attack on the scissle bond of the substrate. [142,151] This generally accepted mechanism is depicted in Figure 3.2 and proceeds as follows.

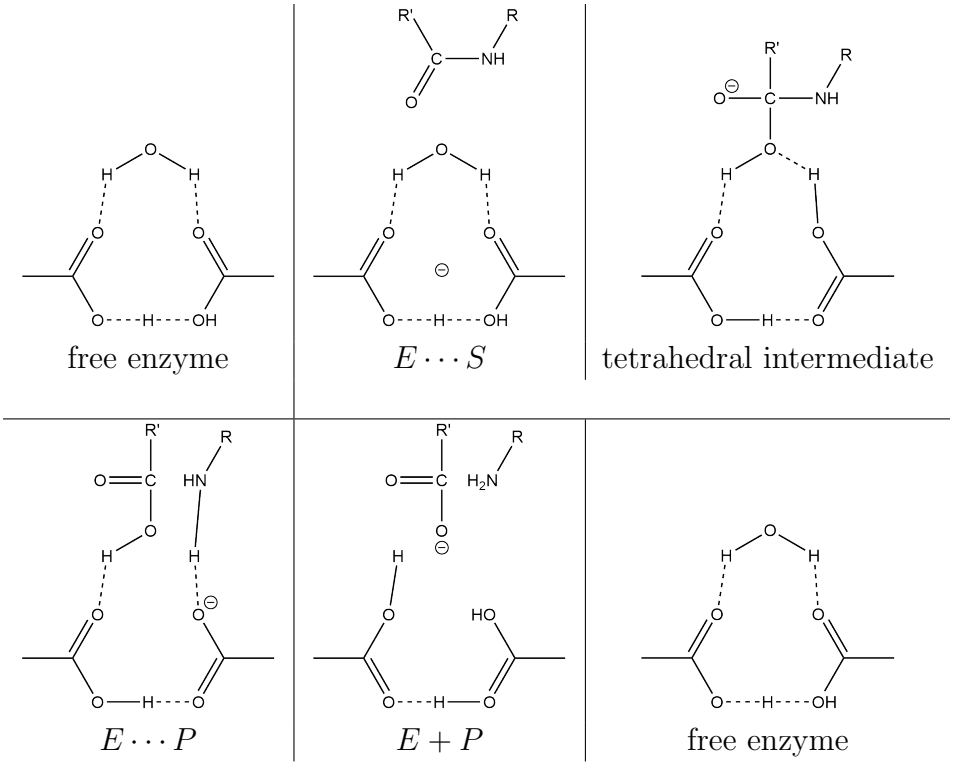


Figure 3.2: Sketch of the catalytic cycle of cleaving peptide bonds. [142,151]

The active site of the free enzyme resides in the 10 membered cyclic structure described for the monoprotonated form of the catalytic dyad above.

Upon binding of the substrate, the carbonyl function of the peptide bond to be cleaved gets positioned in direct vicinity to the oxygen atom of the water molecule, which is bridging the active site residues ASP25/ASP25'. Linked by the low barrier hydrogen bond, the two aspartic acid residues act as base, abstracting a proton from the water molecule which in turn forms a covalent bond via nucleophilic attack onto the carbonyl funtion of the peptide. The degree of the proton transfer is not exactly defined, such that a significant stabilization via the interaction between the oxygen atoms of the water molecule and the transferred proton is conserved. This also prearranges the hydrogen atom in a position, in which it can readily be shifted onto the nitrogen atom of the peptide bond. The last shift allows the formation of a carboxylic acid and an amine function via cleavage of the scissle bond. The reaction products are thus formed and can be released from the active site. The remainder of Figure 3.2 sketches a possible pathway for the release of the products and the restoration of the catalytically active state of the enzyme as seen at the beginning of the catalytic cycle. Some minor details of the mechanism such as the exact sequence of releasing the products or the numerous hydrogen shifts have not yet been fully confirmed, nevertheless the depicted mechanism can be considered as generally accepted.[152]

Depending on the mode of action, the inhibition of HIV-1 Protease can also be described as a simple one- or two step reaction depicted in equation 3.2. For inhibitors that only bind via non-covalent interactions to the target protein, the reaction can be simplified to the first step of the reaction, the formation of the $E \cdots I$ complex out of the separated enzyme (E) and inhibitor (I). In this case, the inhibition potency of the inhibitor can be described solely by the association constant K_i of the non-covalent complex. The second step depicted in equation 3.2 is limited to inhibitors, that act covalently with the target protein, usually by forming a bond with one of the active site residues. Depending on the difference in the free energy of the non-covalent complex ($E \cdots I$) and the covalent adduct ($E - I$), the second step can be highly exergonic ($k_{\text{forward}} >> k_{\text{back}}$) and thus irreversible, energetically neutral ($k_{\text{forward}} \approx k_{\text{back}}$) or even endergonic ($k_{\text{forward}} < k_{\text{back}}$), in which the formation of a covalent bond is very unlikely.

$$E + I \overset{K_i}{\rightleftharpoons} E \cdots I \underset{k_{\text{back}}}{\overset{k_{\text{forward}}}{\rightleftharpoons}} E - I \tag{3.2}$$

The prerequisites for a successful covalent inhibitor as thus a high binding affinity (K_i) towards the target enzyme, such that the formation of the $E \cdots I$ complex can readily occur and a favorable reaction free energy of the bond formation, resulting in the equilibrium between the non-covalent and the covalent complex being shifted towards the later one. For the bond formation, the mechanism has to differ from the catalytic one, as a bond between enzyme and inhibitor and target protein can only be formed, if the

nucleophilic attack happens in a direct and not a water mediated manner. This mode of action was found for the 1,2-epoxy-3-(p-nitrophenoxy)-propane (EPNP) molecule, which is known to be a general, non-specific inhibitor for aspartic proteases[153] and thus also inhibits HIV-1 Protease.[154,155] The nucleophilic attack on one of the epoxides carbon atoms leads to a ring opening reaction, liberating an alcoholate moiety which finally gets protonated by the second aspartic acid residue (see Figure 3.3). Whereas the direct attack is supported by experimental results, the way of the proton transfer reaction is not fully clarified yet. Most probably, the protonation can occur in a direct fashion or mediated by a single molecule of water, bridging the space between the carboxylic acid an the emerging alcoholate.[92,93]

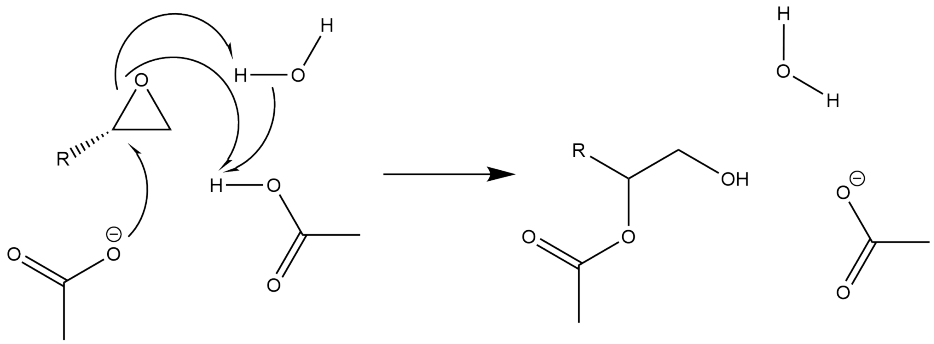


Figure 3.3: Proposed reaction mechanisms for the inhibition of HIV-1 Protease by epoxides such as EPNP. While the nucleophilic attack of the carboxylate onto the epoxide is assured, the proton transfer reaction might occur directly or alternatively mediated by a water molecule.

The formation of the covalent enzyme-inhibitor complex of HIV-1 Protease with EPNP is largely determined by the energetically favorable ring opening reaction of the epoxide which releases the ring strain of the three membered ring and increases the exothermicity of the inhibition reaction. Hence, the epoxide function represents a good lead structure in the rational design of covalent drugs.

3.1.4 Protonation State Dependence on Ligands

During the later stages of the investigations, the generally assumed protonation state of the free enzyme did no longer seem to be valid upon binding of inhibitors. There have indeed been report about changes of the protonation state due to the presence of ligands. Whereas many inhibitors are constructed such to provide both a proton donor as well a an acceptor to form at least two

hydrogen bridges with the ASP25/ASP25' residues,[89,143,156] there are also several reports on significantly different binding modes. A common building block for inhibitors is a vicinal diol, being able to provide two protons, each one being bound to the oxygen atoms of either of the catalytic aspartates. Hence, both aspartates have to reside in their carboxylate form, such that they can act as hydrogen bond acceptors. Similarly, protonated pyrroles where investigated as potential scaffolds which can also provide a twofold hydrogen bond donor functionality, again causing the protein to reside as two carboxylates.[100,109,145,146,157,158,159] According to these findings, the investigations are concentrated on the monoprotonated form of the protein while taking other protonation states into account when required.

3.2 State of the Art

Several preliminary studies have been carried out previous to the current investigations. Searching for a suitable crystal structure of HIV-1 Protease complexed with the known EPNP delivered no result. Instead, a crystal structure of the Simmian Immunodeficiency Virus (SIV) Protease with the desired inhibitor could be found (pdb ID 2sam[113]). The SIV Protease is closely related to the HIV-1 Protease, exhibiting an identical composition of two 99 amino acid residues long protein strands and identical active site substructures. Furthermore, the tertiary structure of the two enzymes is almost identical as can be seen from the overlay in Figure 3.1. In addition to that, a crystal structure of HIV-1 Protease complexed with a different inhibitor from the same group (pdbID 1yth[112]) could be obtained. Hence, a hybrid of the two crystal structures could be constructed by superimposing the active site regions and building a model system of the HIV-1 Protease protein strand in the covalent complex with EPNP. This model system is expected to represent the real enzyme-inhibitor complex sufficiently well for the ongoing investigations. The first series of preliminary calculations used small model systems derived from the constructed covalent complex to study the thermodynamics of the inhibition reaction, the most probable protonation path and the influence of modifications of the inhibitor model.[160] Besides changes in the substitution pattern of the functional group, also an exchange of the epoxide against its nitrogen analog (AZNP) was studied. This modification was made as the nitrogen atom provides the possibility to introduce an additional substituent which in the later stages of the design process could be used to improve the affinity of the inhibitor to the target protein.

Further investigations were then carried out using model systems of the whole protein-inhibitor complexes.[161] These combined quantum mechanical / molecular mechanical calculations were used to determine the thermodynamics of the inhibition reaction under consideration of the protein envi-

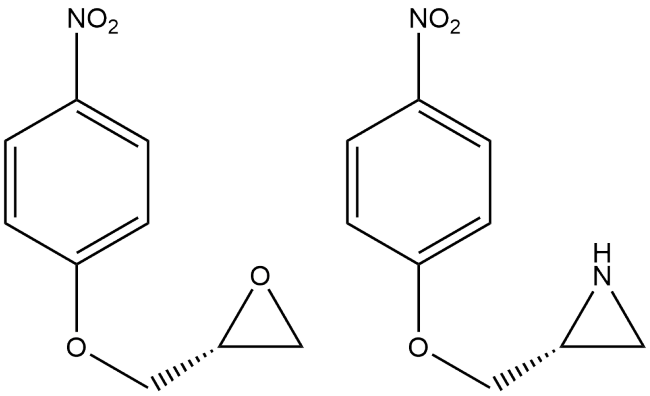
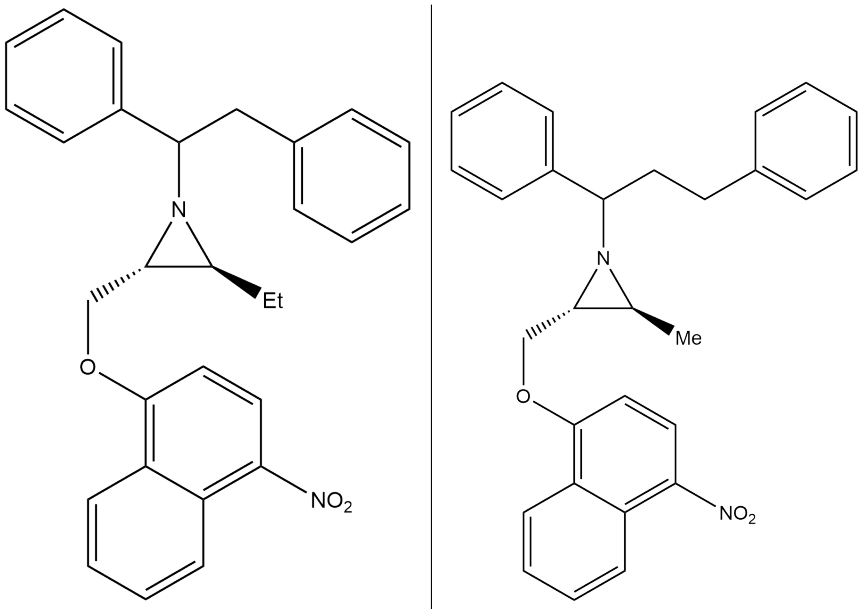


Figure 3.4: Chemical structures of the lead compound, 1,2-epoxy-3-(p-nitrophenoxy)-propane (EPNP) and its aziridine analog, AZNP

ronment, to further clarify potential reaction pathways and to obtain the non-covalent enzyme-inhibitor complexes preceding the irreversible blocked product. These calculations were carried out based on the covalent complex constructed as hybrid structure described above. Whereas these computations also concentrated on the EPNP molecule and its aziridine analog, the employed methods allowed to take environmental influences of the protein surrounding into account. The calculations on the inhibition reaction resulted in comparable reaction energies as those for the pure quantum mechanical ones before. In addition, they delivered the geometrical structures of the preceding non-covalent complexes. These geometries were then used as a basis for docking investigations improving the affinity of the compound. The docking procedures were guided by the search of a substitution pattern which addresses the mostly hydrophobic binding pockets adjacent to the catalytic center. The two most promising candidates are the two compounds depicted in Figure 3.5. The modifications comprise the extension of the former nitrophenoxy-group to a naphtalene scaffold and the addition of a small aliphatic tail to the unsubstituted carbon atom of the ring. The most important change however is the exchange of the epoxides oxygen atom for a nitrogen center, which allowed the addition of the two benzene rings being connected by a flexible alkyl linker chain.

The two newly developed compounds exhibited a similar binding orientation as EPNP with a binding mode percentage over 50%. The estimated binding affinities are very similar with 6.79 in the case of DPPA1 and 6.70 for DPPA2. Compared to the calculated pK_i of AZNP (3.66), this represents already an increase of the affinity by a factor of 1,000.



DPPA1 DPPA2

Figure 3.5: Chemical formulae of the two most promising compounds derived directly from the non-covalent complex of EPNP/AZNP via docking procedures.

3.3 Protonation State of the Free Enzyme

In order to validate the model system for the target enzyme, a series of molecular dynamics simulations was carried out on the unliganded protein under consideration of the various possible protonation states of the catalytic dyad (ASP25/ASP25') and the two His69/His69' residues. For each possible combination of protonation states, neglecting doublettes due to the c_2-symmetry of the enzyme, a separate simulation was performed. Although simulating the free enzyme, the initial geometry was taken from the same Xray crystal structure that was also used for calculations on any of the protein-inhibitor complexes (1yth[112]). To facilitate the preparation of the system, the data taken from the RCSB Protein Data Bank[162] was run through the web-based interface of Charmm-GUI[163] in order to obtain files compatible with the Charmm force field.[27] The further preparation of the model systems was carried out using VMD[28] and the autopsfgen plugin. The later required setting up the correct protonation state for the four amino acid residues of interest, adding the hydrogen atoms missing in the Xray dataset, solvating the protein in a spherical shell of water with a radius of about 50Å and the generation of the corresponding input files for the actual calculations. The calculations were carried out using NAMD[29] as a series of constrained and unconstrained minimizations, a heating procedure, equilibrations and the final molecular dynamics simulations. The initial minimizations were carried out by fixing all heavy atom at their initial coordinates an only relaxing the hydrogen atoms. This step was performed to remove sterical stress caused by the addition of hydrogen atoms to the crystal structure geometry. By only optimizing the hydrogen positions, deviations of the protein structure were reduced to a minimum. The relaxation of the hydrogen atoms was followed by a 10,000 steps long relaxation of the whole system to further reduce artificial stress. The optimized systems were successively heated from 0 to 310K. The temperature range was separated into steps of 10K each and a short simulation of 5,000 timesteps was performed at each temperature step. The heated system was then further subjected to an equilibration phase of 30,000 timesteps to obtain realistically distributed velocities for each individual nucleus before carrying out the actual molecular dynamics simulation. During the equilibration and simulation, no constraints were used on individual atoms. Nevertheless, a spherically symmetric, harmonic potential starting at a radius of 56.5Å was employed to avoid single molecules of water from diffusing away from the system. The minimal radius of that potential was chosen such that the system can evolve in an unbiased way, as usually none of the water molecules reached that outer layer.

With respect to the overall behavior of the free enzyme, all molecular dynamics simulations deliver very consistent results. Depicted in Figures 3.6 are RMSD values measured for each pair of geometries along each of the 18

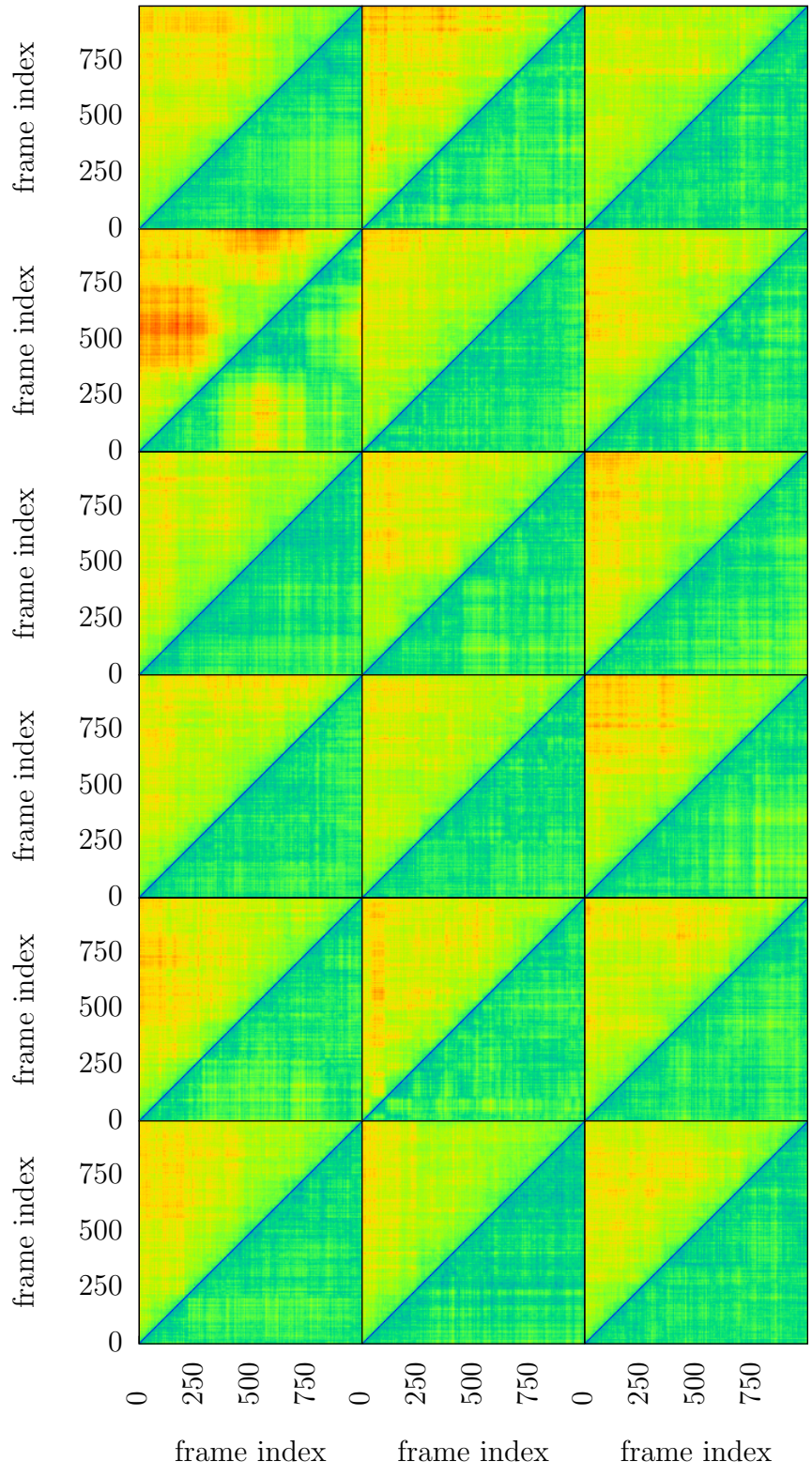


Figure 3.6: RMSD values for the simulation on the free HIV-1 Protease. The alignment was performed on the protein backbone atoms, the measurement on all protein atoms. Range: 0.0(blue) to 3.0Å(red).

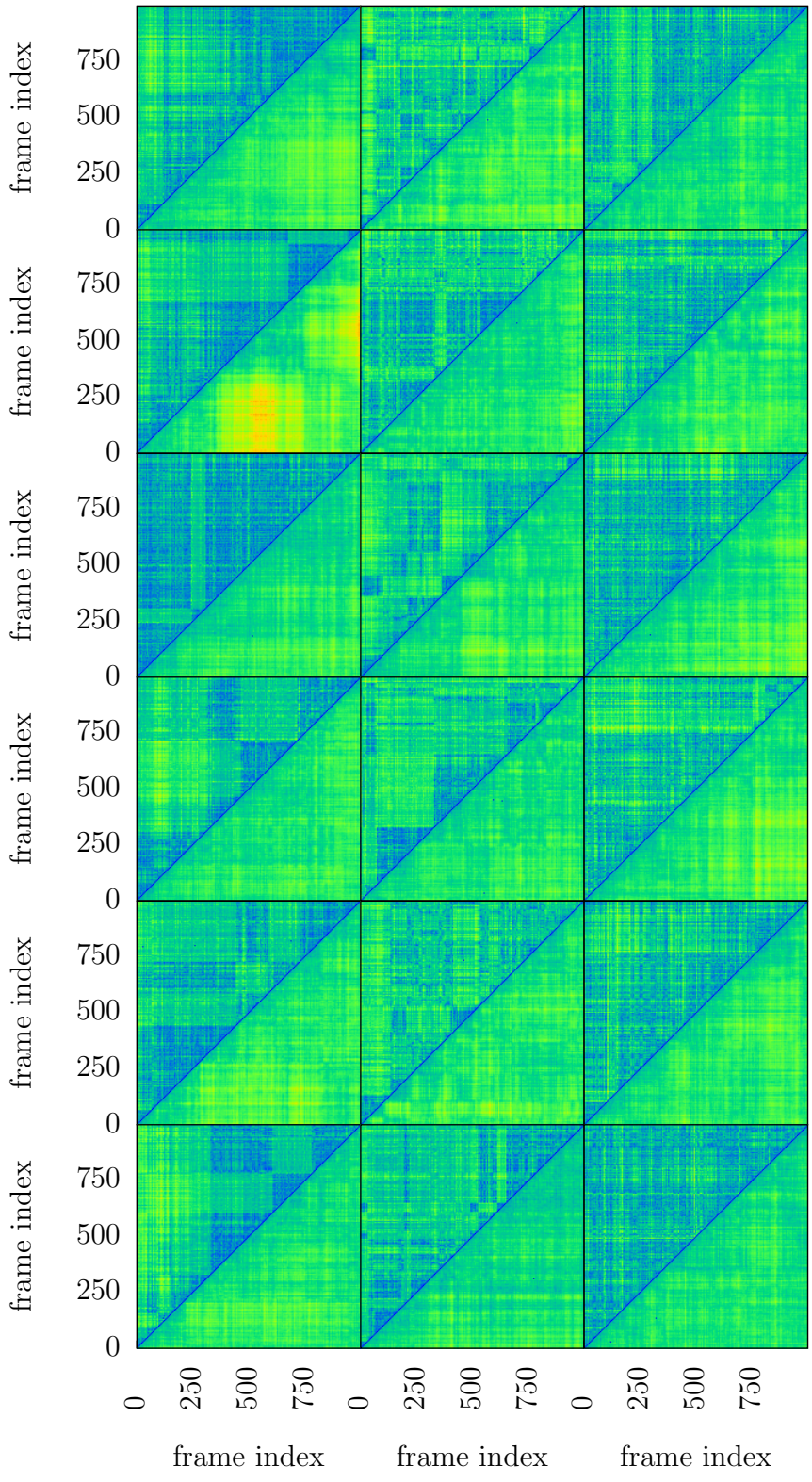


Figure 3.7: RMSD values for the simulation on the free HIV-1 Protease. The alignment was performed on the protein backbone atoms, the measurement on the ASP25/ASP25' residues. Range: 0.0(blue) to 3.0Å(red).

individual trajectories. Each column represents one of the three protonation states of the catalytic dyad, from left to right: deprotonated, mono- and doubly protonated. The protonation states of the His69 residue are, from top to bottom, cationic (first three lines) and protonated at the δ- (4th and 5th line) or ϵ-nitrogen center. Correspondingly the protonation states of the His69' are cationic (1st line), δ-protonated (lines two and four) and ϵ-protonated (remaining lines). The lower right half of each graph represents the deviation of the heavy atoms forming the protein backbone (N, C_α, C, O) while the upper left half provides the values measure for all heavy atoms of the protein, including the side chains of each amino acid residue. As expected from the comparison of crystal structures in Figure 3.1, the deviations of the protein backbone are extremely small, amounting to less than 2.4Å throughout all simulations. The same can be found regarding also the more flexible side chains. Including them into the RMSD measurements, the maximal value remains still well below 3.0Å. Partially unexpected are the values pictured in Figure 3.7. These measurement have been performed for the two catalytic Asp25/Asp25' residues, for which a structural fluctuation below that of the protein backbone was found. With the tertiary structure of the protein being already extraordinarily stable, RMSD values of less than 2.3Å indicate a very rigid geometry of the active site region.

To get a deeper insight into the influence of the protonation states of the ASP25/Asp25' dyad as well as that of the His69/His69' residues, distance measurements between the C_γ atoms of the catalytic dyad were performed. Whereas the protonation state of either histidine did not have any noticeable effect on the protein geometry, that of the catalytic dyad could clearly be seen. Consistently throughout all simulations, in which the two Asp25/Asp25' residues had been treated as negatively charged carboxylates, their electrostatic interaction led to significantly increased distances between their C_γ atoms (5.16Å) compared to either the monoprotonated (4.00 Å) or fully protonated systems (4.26 Å). The additional hydrogen atom in the monoprotonated state does not only reduce the repelling electrostatic force between the ASP25/Asp25' residues, but it also allows the formation of a hydrogen bond between them, which provides further stabilization. A visual inspection of the geometries along the trajectories of the corresponding simulations also revealed, that the assumed cyclic structure (see Figure 3.2) can frequently be encountered. Rather unexpected is the fact, that the distances measured for the monoprotonated systems are even lower than those of the systems comprising a fully protonated catalytic dyad. With both catalytic residues being protonated, repelling forces due to negative charges should no longer exist and the presence of two carboxylic acid moieties should result in the formation of very stable dimers, held together by two hydrogen bridges. Nevertheless, the formation of such dimers is hampered in the actual model systems, as the rigidity of the active site region disfavors the two carboxylic

acid functions to directly face each other.

The measured C_γ-C_γ distance were also compared to those determined from five Xray crystal structures. These structures are the one, the protein structure was taken from for building the model system (pdbID 1yth[112]), a structure comprising the free enzyme (pdbID 1hhp[134]) and structures of the free enzyme containing several mutations (pdb IDs 2g69[133], 3kt2 and 3kt5[132]).

simulations		crystal structures	
deprotonated	4.29-7.17Å	1yth	5.02Å
monoprotonated	3.31-5.29Å	1hhp	5.36Å
fully protonated	3.49-5.88Å	2g69	4.94Å
		3kt2	4.90Å
		3kt5	4.97Å

Table 3.2: Distances between the C_γ carbon atoms of Asp25/Asp25' calculated for the molecular dynamics simulations of the free HIV-1 Protease and reference values taken from different Xray crystal structures.

3.4 Inhibitor Models of the first Generation

The docking experiments conducted to improve the binding affinity of the non-covalent complex of AZNP delivered several promising candidates as those depicted in Figure 3.5. From the newly generated molecules, I have chosen DPPA1 as most promising candidate, as it exhibited the larges binding affinity (pK_i=6.79) and the orientation of the reactive aziridine ring resembled the one of AZNP most closely. As the methods used for docking are well established, the generated structures appear highly reliable. However, due to the explosion of the conformational space, the docking experiments had been performed on a rigid protein geometry only. Hence, the obtained structure had to be further refined to account for the flexibility of the protein, especially the exact geometry of the active site and the position of the two flaps. As the refinement was also intended as preliminary step preparing the system for later combined quantum mechanical / molecular mechanical calculations on the bond formation process, the same force field was used to carry out molecular dynamics simulations. The parameterization of the protein and solvent shell was done identically to the simulations of the free protein before. The missing data for the new inhibitors was generated using the web-based interface of Swissparam[164] to produce a valid topology and parameter set to be used with Charmm[27]. As the atomic charges assigned

by Swissparam are known to be of low quality, they have been replaced by those obtained from quantum mechanical calculations on the isolated inhibitor molecule according to the recommendation given in the extension to the Charmm force field, the Charmm General Force Field (CGenFF).[165] The further preparation of the molecular dynamics simulations was carried out similarly to that used for the free enzyme before. The water shell however was not newly generated, but the already optimized one from the preceding calculations was used. As the docking experiments did not yield a single optimal orientation of the inhibitor but several equally favorable ones, eight separate simulations for each of them were conducted. The analysis of these simulations was started by an analysis of the overall stability of the protein structure and the flexibility of the catalytic dyad as described before. Again, the simulations confirm a very rigid tertiary structure of the enzyme, for which the RMSD values calculated for the atomic positions of the protein backbone atoms (N, C_α, C, O) did not exceed 2.3Å which is comparable to the value obtained for the free enzyme (2.4Å). Similarly the fluctuations of the overall geometry remain very low with RMSD values of less than 2.9Å throughout all simulations (see Figures 3.8). Looking more closely on the catalytic residues, the two Asp25/Asp25' also exhibit an extremely stable conformation with their RMSD values being well below 2.2Å (see Figure 3.9).

In addition to measurements on the protein atoms only, also the fluctuations in the position of the inhibitor could be analyzed. For the inhibitor measurements had been performed on the whole molecule, represented by its heavy atoms only, and the reactive functional group, comprising the three atoms of the aziridine ring. A first remarkable result is that the inhibitor model as a whole exhibits a flexibility almost twice as large (up to 4.98Å) as that of the surrounding protein, indicating larger structural changes in the orientation of the inhibitor. Even more surprising are the RMSD values determined only for the three rings of the aziridine ring, for which the measurements yielded up to 5.43Å. A visual inspection of the molecular dynamics trajectories revealed that although addressing the hydrophobic pockets of the target protein, the phenyl rings do not completely fill the hydrophobic binding pockets of the HIV-1 Protease, thus retaining a notable flexibility. The reactive aziridine group however does hardly show specific interactions with the protein surrounding, leaving even more flexibility both to its position within the active site as well as to is orientation.

While the newly introduced substituents help to constrain the molecule, the missing interactions between the aziridine group and the neighborhood of the catalytic dyad results in a high tendency for the electrophilic group to moved away from the Asp25/Asp25' residues. This is depicted for the simulation of the best ranked docking pose by the evolution of the interatomic distances between the potentially nucleophilic oxygen atoms of Asp25 and

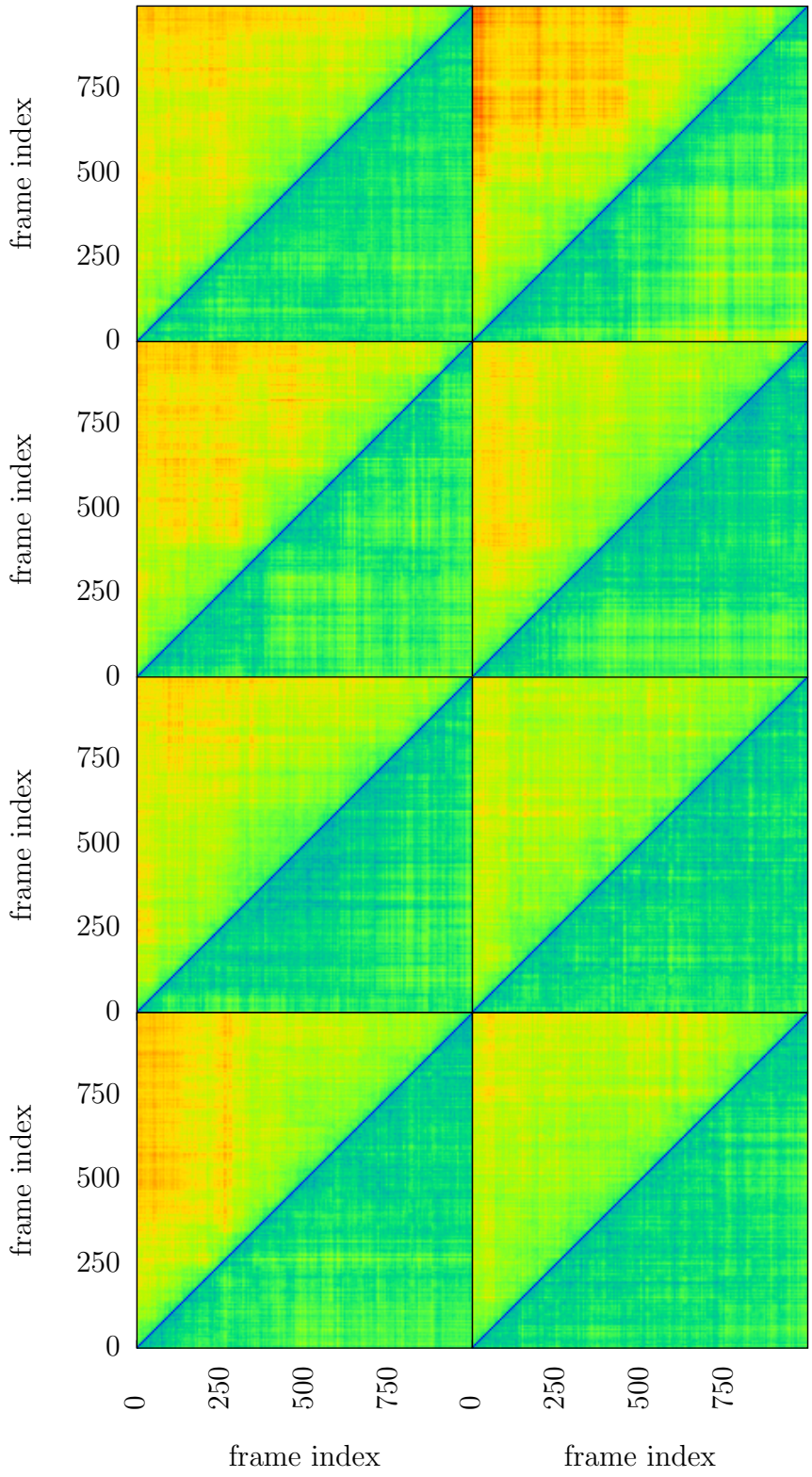


Figure 3.8: RMSD values for the simulations on HIV-1 Protease in complex with DPPA1. The alignment was performed on the protein backbone atoms, the measurement on all protein atoms. Range: 0.0(blue) to 3.0Å(red).

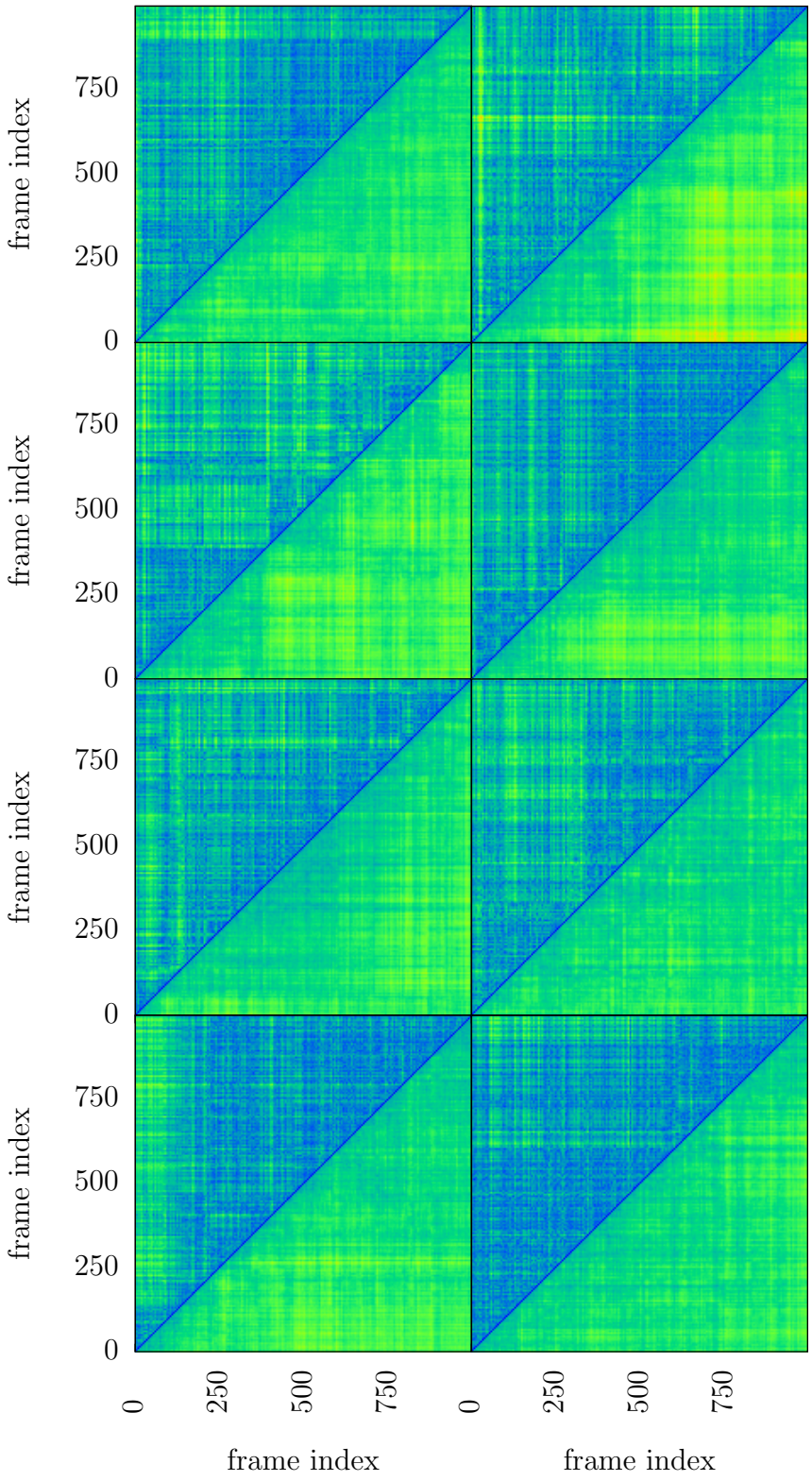


Figure 3.9: RMSD values for the simulations on HIV-1 Protease in complex with DPPA1. The alignment was performed on the protein backbone atoms, the measurement on the catalytically active Asp25/Asp25' residues. Range: 0.0(blue) to 3.0Å(red).

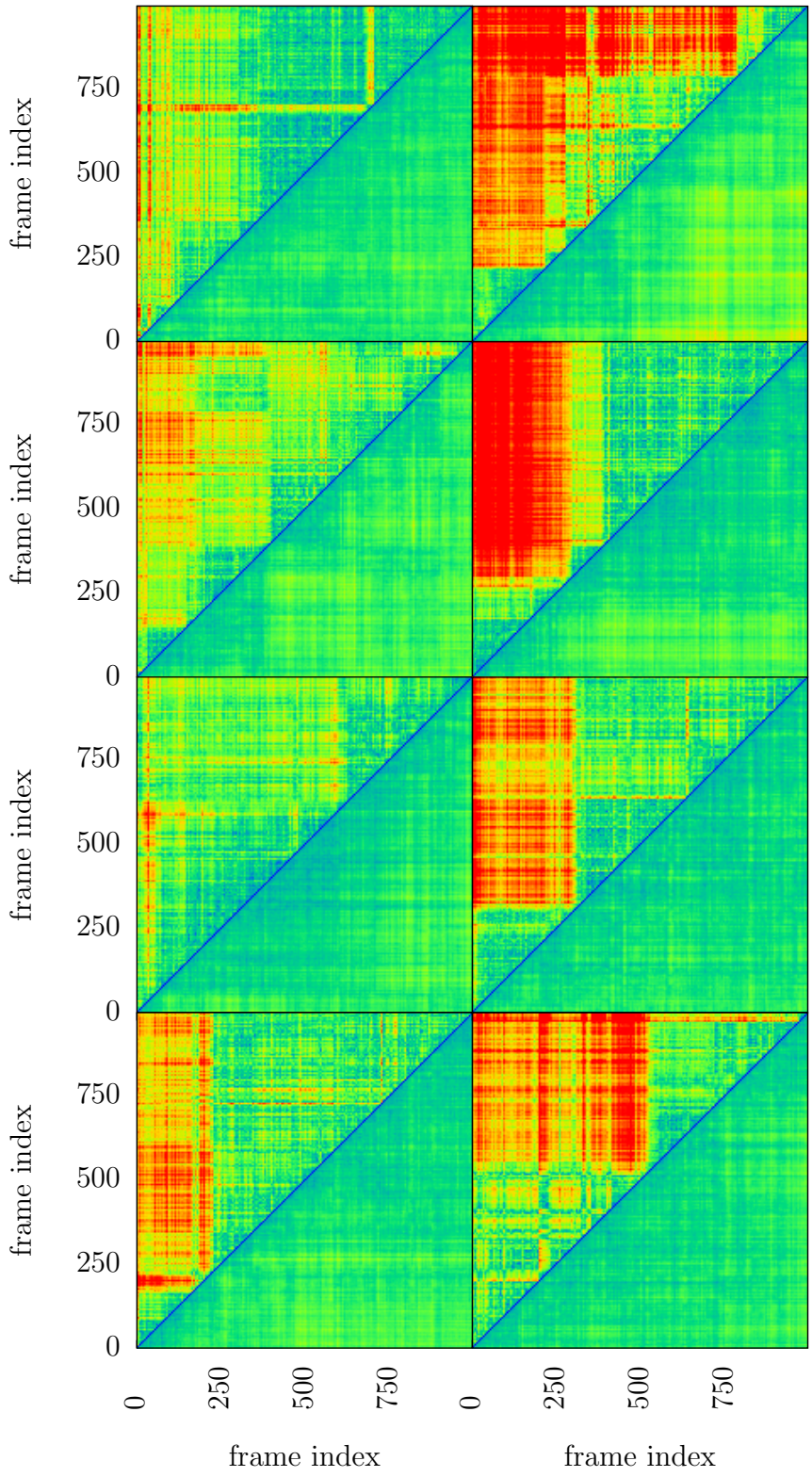


Figure 3.10: RMSD values for the simulations on HIV-1 Protease in complex with DPPA1. The alignment was performed on the protein backbone atoms, the measurement on all non-hydrogen atoms of the inhibitor. Range: 0.0(blue) to 3.0Å(red).

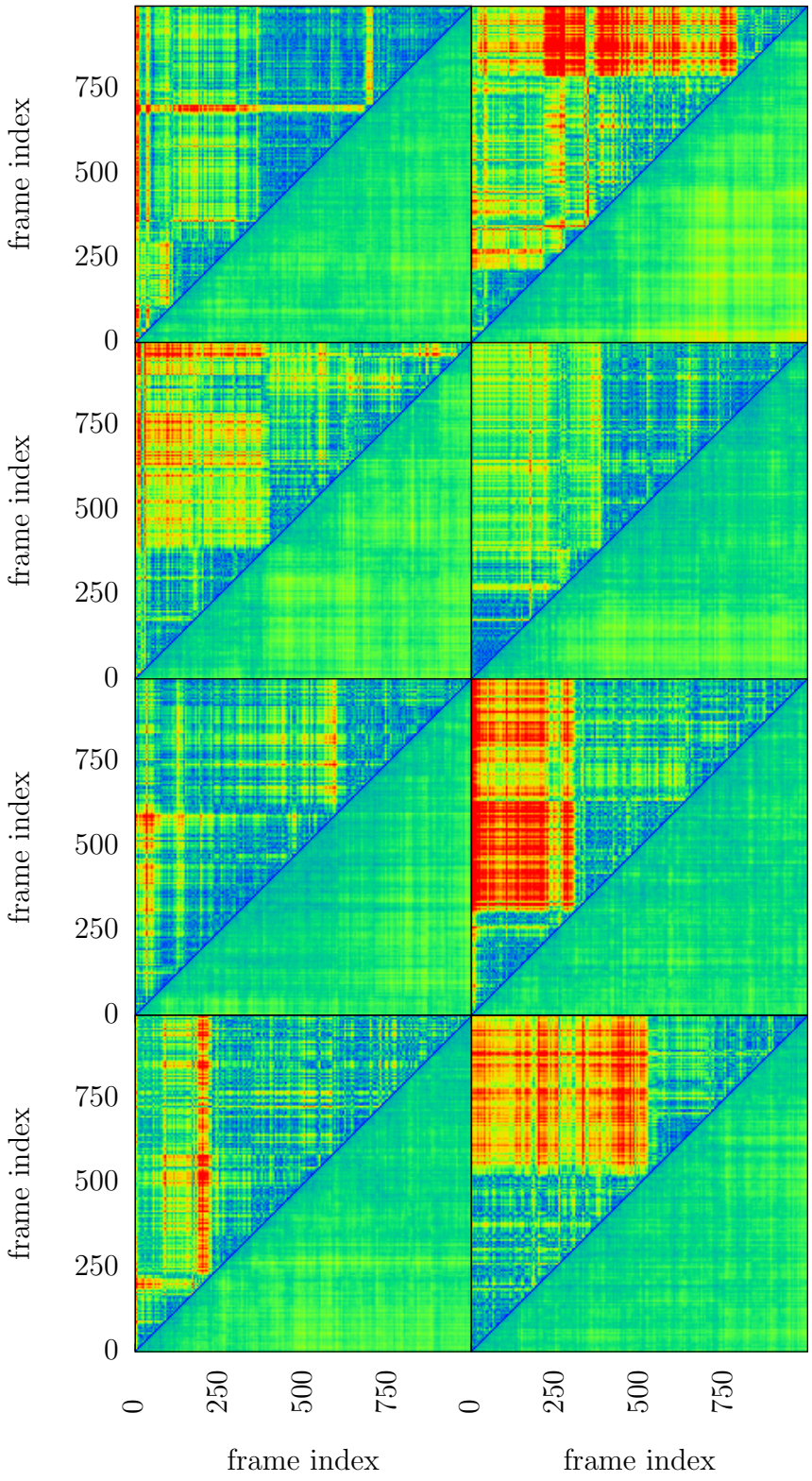


Figure 3.11: RMSD values for the simulations on HIV-1 Protease in complex with DPPA1. The alignment was performed on the protein backbone atoms, the measurement on the three non-hydrogen atoms of the aziridine ring. Range: 0.0(blue) to 3.0Å(red).

the closer of the two carbon atoms of the aziridine ring (see Figure 3.12. While the aziridine ring remains in a reasonable distance from the catalytic residues during the first quarter of the simulation, a significant increase of the measured distances can be seen at around 300ps. This increase is not only some temporarily constrained fluctuation, but the larger distances remain at the elevated values during the remainder of the simulation. Although depicted only for a single simulation here, the trend is the same for the other simulations with the only notable differences being the exact time when the structural change happens. Nevertheless, as can be seen from Figures 3.10 and 3.11, the structural change happens at the latest after half the simulation period, thus within a frame of 2.5ns.

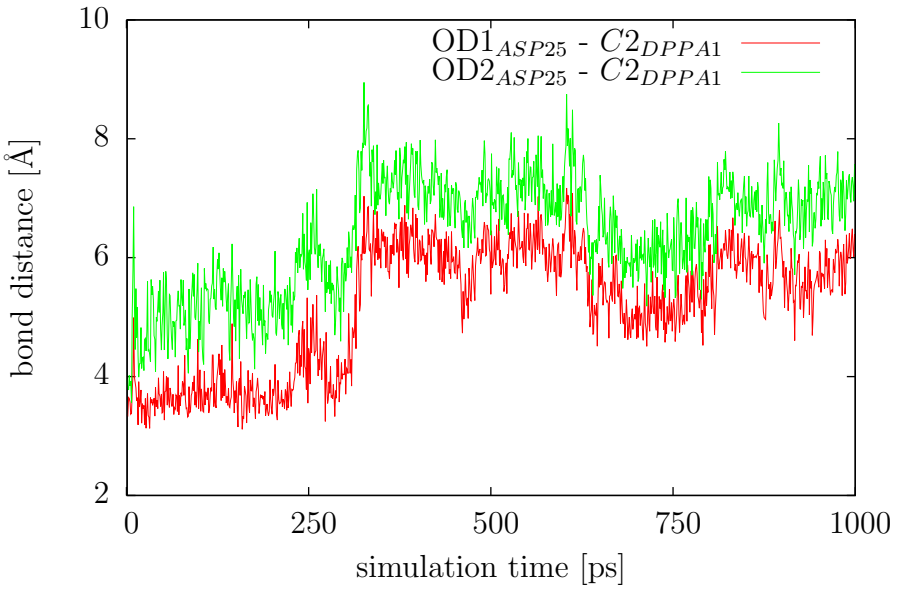


Figure 3.12: Bond distances between the nucleophilic oxygen atoms of the Asp25 carboxylate function and the electrophilic C2 carbon atom of the inhibitor along the trajectory of one of the simulations.

The observed diffusion of the aziridine ring away from the catalytic dyad implies that the necessary non-covalent complex preceding the formation of a covalent bond between the enzyme and the inhibitor with a proper orientation of the aziridine ring with respect to the nucleophilic Asp25 is very unlikely to form. Thus, although the binding affinities of DPPA1 and DPPA2 exceed by far those of the initial compound, a covalent inhibition is not possible.

3.5 Inhibitor Models of the second Generation

In order to resolve the issues observed during the simulations of the first generation of inhibitors, further optimizations of the chemical structure had been made. To benefit from the favorable binding modes of the two aromatic rings introduced as substituents at the aziridine nitrogen center, this part of the molecule was hardly changed with respect to DPPA1. A carboxylic acid function was added to one of the rings, as this part of the molecule remains solvent exposed even in the bound state. Thus adding this carboxylic acid function retains favorable interactions with the surrounding water shell, decreasing the energy loss due to the desolvation of the molecule upon binding. Furthermore it is expected, that the carboxylic acid function might help to facilitate the solubility of the molecule, which would allow an easier administration of the drug later on. For some of the new compounds, an additional CH_2 function was introduced between the substituent and the nitrogen atom. The expected advantage of this modification is an increased flexibility of the nitrogen center, which is mandatory for the ring opening reaction, while the aromatic rings are allowed to remain at their positions most perfectly addressing the corresponding binding pockets. The most significant change between the previous models and the current series is the extension of the aziridine ring to a bicyclic scaffold comprising the three membered aziridine ring augmented by a pyrrolidine ring. The idea behind the construction of this 3,6-diazabicyclo[3.1.0]hexane scaffold was to rigidify the core of the inhibitor, avoiding large scale movements of the reactive group. The later should be positioned in a reactive orientation in close vicinity to the catalytic dyad, anchored by the hydrophobic substituents. The modifications of the inhibitors core also required modifications of the substitution pattern at the two bridgehead carbon atoms, where only hydrogen atoms of a single methyl group were used to minimize sterical hindrance of the nucleophilic attack. Also due to sterical reasons, the connection between these carbon atoms and the naphtalene unit was elongated by another CH_2 group, whereas the naphtalene moiety itself was retained as it addresses the available binding pocked in a very favorable way. Nevertheless, the naphtalene group did not yet completely fill the available space, hence it was enlarged by an aromatic sulfonamide tail, which is know to be recognized well by the HIV-1 Protease, as it was optimized in the development of the FDA approved drug Tipranavir.[166,167] The modifications of the chemical structure are depicted in Figure 3.13.

Among the best ranked docking candidates, the four molecules of Figure 3.13 reached the highest scores. Due to the chemical similarity of the four molecules, the high degree of similarity and the huge computational effort required, the validation of the covalent reactivity against HIV-1 Protease was only performed for the topmost molecule, 1020. The investigations on the re-

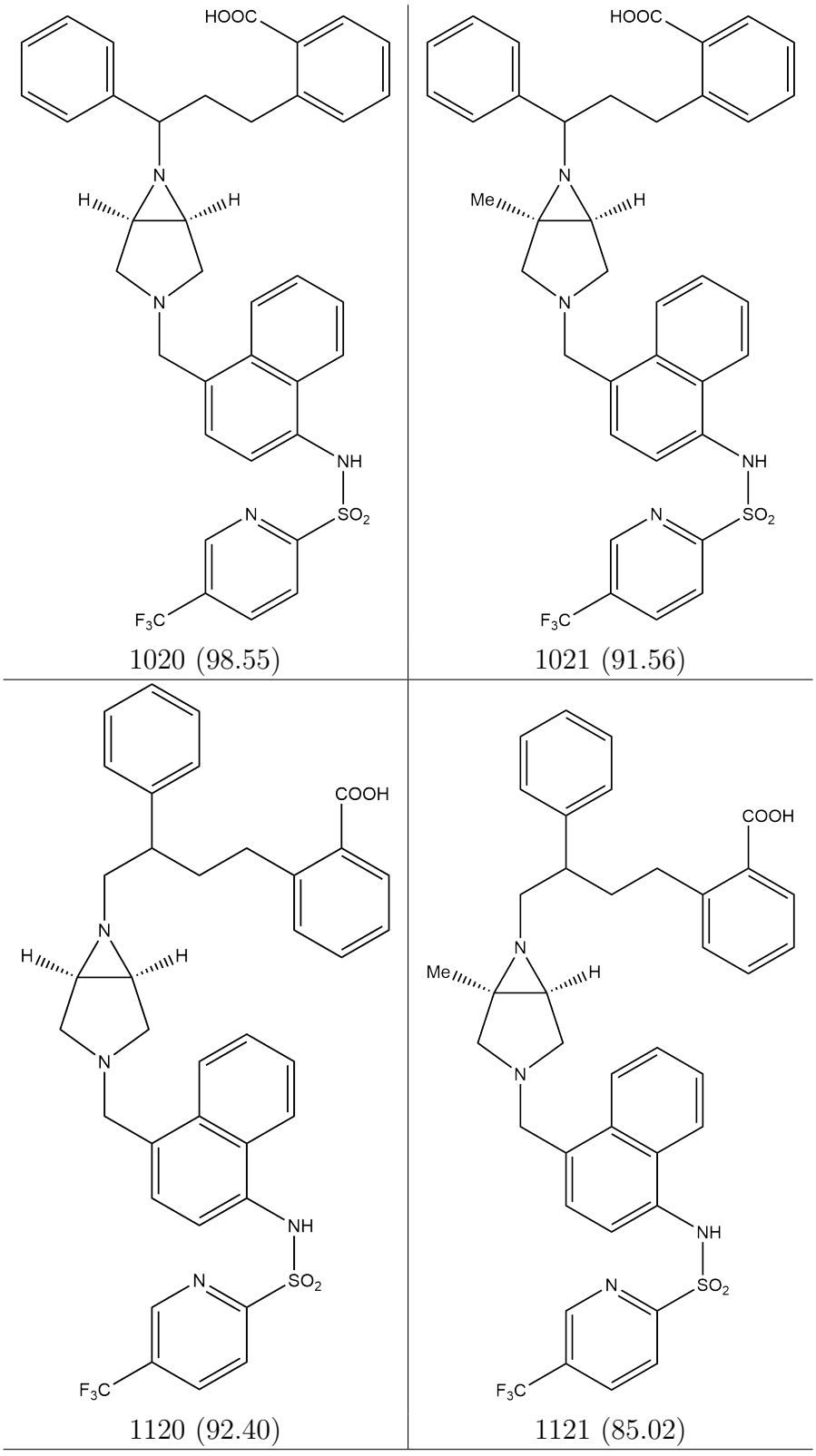


Figure 3.13: Chemical formulae of the inhibitor molecules developed during the second interation of docking. Values provided in parenthesis fitness scores determined with Goldscore.[168,169]

activity of 1020 were started with a series of molecular dynamics simulations on the best ranked inhibitor pose.

3.5.1 Preliminary Molecular Dynamics Simulations

The docking investigations have been carried out on a rigid protein structure and, in order to exploit the c_2-symmetry of the enzyme, with the catalytic Asp25/Asp25' residues being treated in their carboxylate form, such that each of the two amino acid groups could potentially act as nucleophile. Nevertheless, the reaction mechanism does not only involve the nucleophilic attack on one of the two bridgehead carbon atoms of the inhibitor but also the protonation of the nitrogen center, which gets detached from it during the ring opening reaction. The required proton donor is expected to be the second catalytic aspartate, thus for the further investigations, it was treated in its carboxylic acid form. This is also in line with the assumed protonation state of the enzyme providing a carboxylate and a carboxylic acid group. The decision of where to place the missing proton, on the Asp25 or the Asp25' residue, was determined by the shortest distance between one of the carboxylate oxygen atoms and one of the two electrophilic carbon atoms of 1020, hence the most probable path for the bond formation reaction. The proton was thus placed on the other aspartate. For the system, five individual simulations were started using the same initial geometry, but separate heating and equilibration steps before carrying out the actual simulations. The technical details of these preparatory steps as well as the simulation parameters are the same as those described before. The reason for starting several simulations on the same initial geometry is that during the preparation of the MDs, the initial velocities of the individual atoms are assigned randomly, though Boltzmann distributed, velocities. Hence, depending on the starting conditions, the molecular dynamics simulations could evolve in a very diverse manner, increasing the chance of sampling rare events such as diffusion processes of the changes in the position of the inhibitor within the active site.

3.5.2 Selecting Geometries for QM/MM

The scoring function is defined by the number of quantities (distances, angles or dihedral angles) measured for the system of interest. Each measured values r_i is compared to the corresponding optimal value R_{opt} as well as a validity range between the lower r_{min} and upper r_{max} boundary. If the value of r_i i outside the valid range, it will not contribute to the total score *Stexttotal* as can be seen from equation 3.3. Within the range, the score is calculated by the euclidean distance between the value r_i and its optimal value r_{opt}, normalized to the distance between r_{opt} and the respective boundary. Thus,

the individual contributions to the total score can obtain values between zero and one. A weighting factor c_i can be used to account for different importance of the selected quantities. With positive values assigned to c_i, r_i is in favor of the current frame if it is in the provided range. Conversely, negative values of c_i provide a means to assign penalties to geometries, which obtain an arrangement that is known to be disfavored.

$$S_{\text{total}} = \sum_{i=1}^{N_{\text{quantities}}} c_i \times \begin{cases} 0 & , \text{ if } r_i < r_{\min} \\ 1 - \left| \frac{r_i - r_{\min}}{r_{\text{opt}} - r_{\min}} \right| & , \text{ if } r_{\min} \leq r_i \leq r_{\text{opt}} \\ 1 - \left| \frac{r_{\max} - r_i}{r_{\max} - r_{\text{opt}}} \right| & , \text{ if } r_{\text{opt}} < r_i \leq r_{\max} \\ 0 & , \text{ if } r_{\max} < r_i \end{cases} \tag{3.3}$$

In the present case for the simulations of 1020 and the HIV-1 Protease, only interatomic distances have been used as criteria in selecting most useful geometries from the simulation trajectories. Six measures have been performed for the distances between:

- the two carboxylate carbon atoms (C_γ) of the Asp25/Asp25' residues
- the proton of the carboxylic acid function of Asp25 and the two carboxylate oxygen atoms of the Asp25' residue
- the nucleophilic oxygen atoms of Asp25 and the two bridgehead carbon atoms of the inhibitor
- the aziridine nitrogen atom and the proton of the carboxylic acid function of Asp25

The parameters $r_{\min}$, r_{opt} and $r_{\max}$ were chosen such that the optimal C_γ-Cγ distance resembles that measured in the crystal structure of 1yth and that a suitable trajectory for the nucleophilic attack as well as a plausible protonation path between the nitrogen and the carboxylic acid function can be found. The weighting factors chosen such that contributions to the score by equivalent distance measures are not doubly counted. An adjustable threshold was defined, such that only a manageable number of structures could be selected. The finally selected geometries comprise the frame 599 of the third trajectory, frame 250 of trajectory four and frame 815 of the last simulation.

3.5.3 QM/MM Calculations on the Reaction Path

To follow the reaction path by means of combined quantum mechanical / molecular mechanical calculations in form of potential energy surfaces, two

internal degrees of freedom were chosen as representing the major components of the reaction coordinate. The first coordinate was chosen as being the distance between the two C_γ atoms of the Asp25/Asp25' residues. While being hydrogen bonded during the simulation, increasing the distance between them should break up the hydrogen bond, liberating the proton to be transferred to the aziridine nitrogen atom while the other aspartate should increase in nucleophilicity thus facilitating the nucleophilic attack on the inhibitor. The corresponding distance between the attacking oxygen atom and the electrophilic carbon atom of the inhibitor was used as a second coordinate, giving a direct measure of the bond formation process. To reduce the computational demands of the calculations, the water shell compassing the whole system during the simulations was truncated to a small sphere with only 15Å radius around the inhibitor molecule. The active region of the QM/MM system was yet smaller, comprising only residues within a distance of 12Å around the inhibitor. Thus the outermost shell of water molecules fixed at their initial positions provides a reasonable boundary between the center of the chemical reaction and the surrounding solvent while the molecules within this shell provide sufficient flexibility to adopt to the actual position along the reaction path. The reaction path was followed in 0.1Å increment along each of the coordinates described above. All potential energy surfaces exhibit a clear minimum at the position of the initial, non-covalent enzyme-inhibitor complex, indicated a good agreement between the purely molecular mechanical calculations and those on the QM/MM level. Nevertheless, a defined reaction path from the non-covalent to the covalent complex could not be determined for any of the investigate systems. For the first of the potential energy surfaces, the energy exhibits a steep increase with larger distances between the two C_γ atoms, which could be easily explained with the reduction of the stabilization through the hydrogen bond between the two carboxylate functions. Nevertheless, the direction with modest ascend in energy does not lead towards the covalent complex. Instead, the carboxylic acid moiety turns away from the carboxylate, forming a hydrogen bond with the protein backbone beneath the active site and thus making the proton unavailable for a transfer to the inhibitor. For the potential energy surface calculated for the second selected geometry, a similar behavior could be found. Because of the differences in the exact geometry of the surrounding protein structure, especially the more flexible side chains, the potential around the initial minimum is a lot more shallow compared to the first system. Nevertheless, the increase in the Cγ-Cγ distance also leads to the loss of the stabilizing hydrogen bond, which is compensated for by a similar reaction of the carboxylic acid function, that reorients to form another H-bond to the protein. Also in this case, a proton transfer to the inhibitor is no longer possible. Only for the last of the investigated systems, an approximation of the desired reaction path could be found. Visual inspection of the structural changes along the

path revealed, that the protein responds with the relocation of two short protein loops to stabilize the structural changes caused by the separation of the two Asp25/Asp25' carboxylates. These loops comprise the residues Thr26 to Ala28 adjacent to the Asp25' residue and the two Leu24 and Thr26 residues next to Asp25. Also promising is the much smaller increase in energy upon shortening of the distance between nucleophile and inhibitor, although the relative energies do not suggest an exothermic reaction.

Summarizing the difficulties observed throughout the QM/MM potential energy surfaces on the systems described above, the major problem seems to be the position of the proton required for the ring opening reaction. Whereas favorably stabilized between the two carboxylates during the molecular dynamics simulations, there is no favorable path for the proton to be shifted to the inhibitor. Nevertheless, as experienced with the very first quantum mechanical model systems,[160] in the case of aziridine-based inhibitors, the proton transfer has to occur much earlier compared to epoxide-based ones, hence before the ring opening is started. Therefore, the proton transfer could not only be expected to proceed once the inhibitor has completely reached its geometry within the non-covalent complex, but it could also happen during the binding process or even before. This is also strongly supported by a comparison of the low and high pK_a values of the catalytic dyad (3.3±0.1 and 6.8±0.1 respectively)[135,136] and the two nitrogen atoms of the inhibitor. An estimation of the pK_a of the aziridine N-atom yields about 7.9[170] and around 11.27[171] for that of the pyrrolidine ring. According to these pK_a values and considering the pH of the surrounding, at least the pyrrolidine nitrogen center should reside in a protonated form while that of the aziridine is also very likely protonated. The catalytic dyad of the protein could thus reside either in the assumed monoprotonated state or, assuming a proton shift preceding the non-covalent binding, in a completely deprotonated form exhibiting two carboxylates.

3.5.4 Validation of the Protonation State

To validate the assumption of protonation states differing from those expected based on the proposed inhibition mechanism, another series of molecular dynamics simulations was carried out. In four separate simulations, four combinations have been investigated according to Table 3.3. The monoprotonated protein complexed with the neutral inhibitor (NM) corresponds to the system in the originally expected state. According to the pK_a values, adding a single proton to the inhibitor, the favored position should be the pyrollidine nitrogen center, resulting in the PM system. The other two systems were constructed by completely deprotonating the catalytic dyad and using the inhibitor in its singly (PD) or doubly protonated (DD) form.

The focus of the analysis of these four simulations was the degree of dimer-

protein	inhibitor	label
monoprotonated	neutral molecule	NM
monoprotonated	protonated at pyrollidine	PM
deprotonated	protonated at pyrollidine	PD
deprotonated	protonated at both N-centers	DD

Table 3.3: List of combined protonation states of the non-covalent complexes between HIV-1 Protease and the 1020 inhibitor molecule.

ization between the catalytic aspartates, the position and orientation of the electrophilic carbon atoms of the inhibitor facilitating the formation of a covalent bond, the structural stability of the overall structure as well as the comparison between the simulated structures with available Xray crystal structures. RMSD measurements on the protein backbone, the overall protein structure and that of the inhibitor are depicted in Figures 3.14 and 3.15.

As expected, the overall structure of the protein again remains very stable throughout all simulations with RMSD values of its backbone below 2.1Å and those of the whole protein structure not exceeding 2.7Å with only little differences between the individual simulations. In contrast to that, the values determined for the two catalytic aspartates show clear differences between the simulations, at their side chains react to the change in their protonation state. With the protonation state resembling that of the previous simulations, the structural fluctuations are much lower (below 2.0Å) as for those simulations, in which the catalytic dyad was modeled in its deprotonated state. With the missing hydrogen bond and the electrostatic force between the two carboxylates, the side chains of the Asp25/Asp25' residues tend to avoid each other, which is reflected in an increase both in the distance between them as well as in the RMSD values along the trajectory, amounting to more than 2.7Å in the case of the PD system. The hydrogen bond between the two catalytic residues leads to a dimerization between them, forcing the inhibitor away from the Asp25/Asp25' dyad and towards the flaps, making the formation of a covalent bond less probable. For the systems with the deprotonated catalytic dyad however, the inhibitor frequently adopts a favorable position for the bond formation, for which in this case either of the two carboxylates could act as a nucleophile. During all simulations, water molecules from the solvent tend to move close to either the carboxylates of either of the protonated nitrogen atoms of the inhibitor, frequently forming hydrogen bonds between them. These hydrogen bonds could either be regarded as stabilizing factors for the formation of the corresponding complexes or as potential proton transfer pathways between the participating

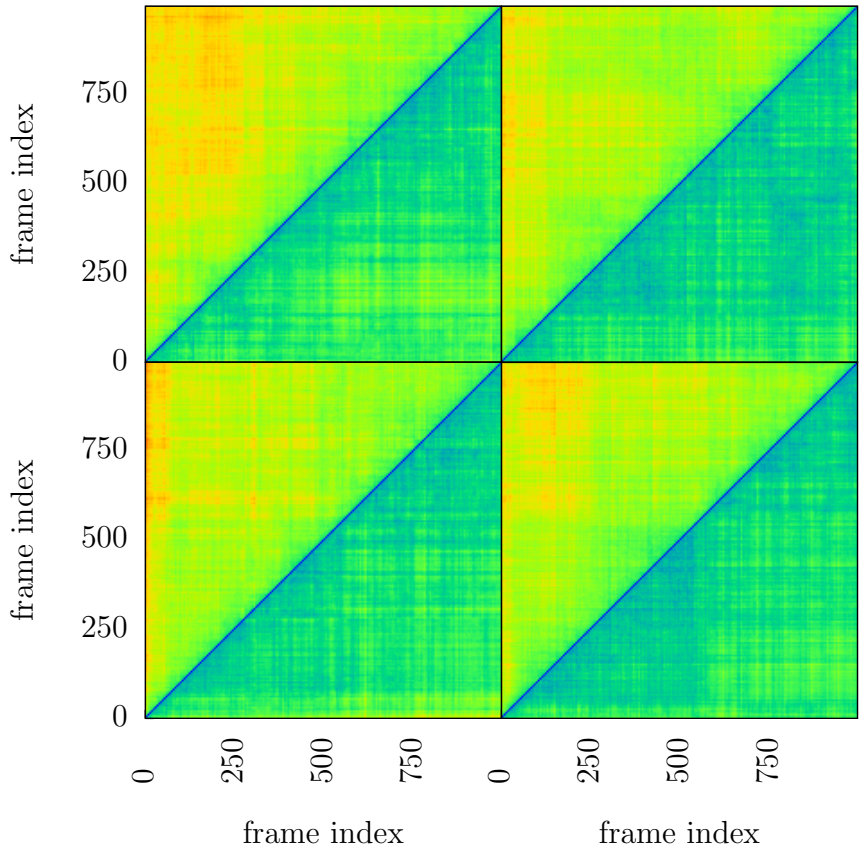


Figure 3.14: Two dimensional Measurement of RMSD values between each pair of frames along the molecular dynamics simulations performed for the alternative protonation states of HIV-1 Protease complexed with 1020. The lower right half represents the geometrical fluctuations of the protein backbone, the other half those of the whole protein. The color range is blue (0.0Å) to red (3.0Å).

heteroatoms.

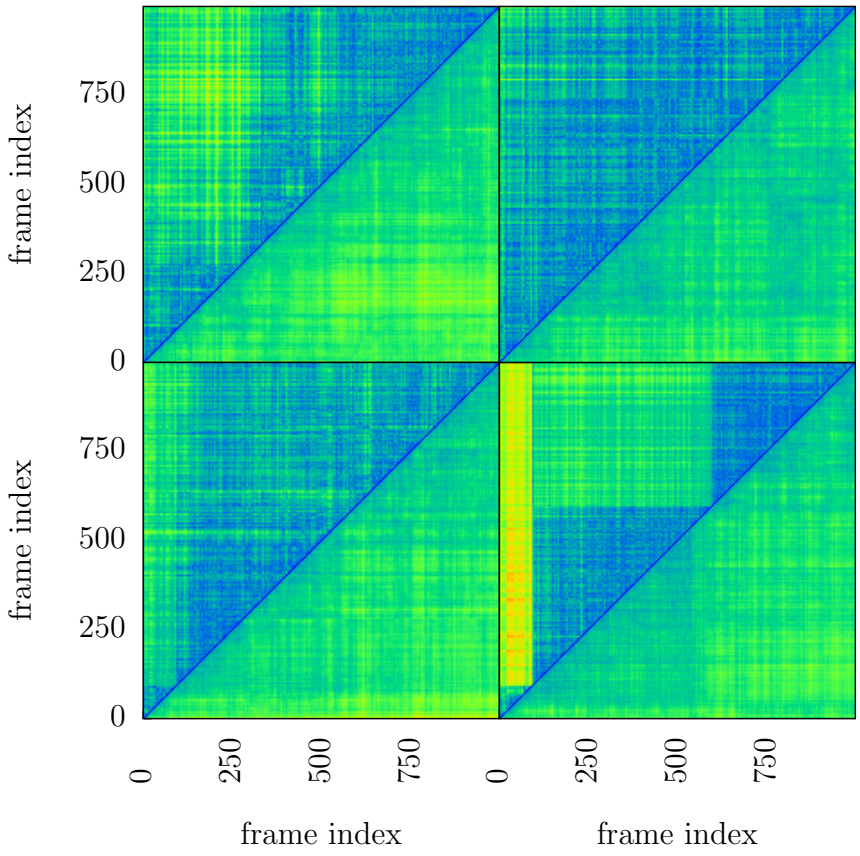


Figure 3.15: Two dimensional Measurement of RMSD values between each pair of frames along the molecular dynamics simulations performed for the alternative protonation states of HIV-1 Protease complexed with 1020. The lower right half represents the geometrical fluctuations of the protein backbone, the other half those of the catalytic residues. The color range is bue (0.0Å) to red (3.0Å).

3.6 QM/MM Calculations on further Protonation States

As the distance measures along the trajectories of the molecular dynamics simulations revealed, a bond formation can only happen in the system where the catalytic dyad exists in form of two carboxylates. Therefore only those

systems comprising a twofold carboxylate within the active site have been considered fur the reaction path calculations. The two resulting systems contain the inhibitor molecule either being protonated at its pyrrolidine nitrogen center or being fully protonated carrying a proton on each of the pyrrolidine and the aziridine N-atom. The selection of suitable frames along the trajectories was again guided by several geometrical parameters such as a close proximity between one of the nucleophilic oxygen atoms of Asp25/Asp25' and either of the two electrophilic carbon atoms of the inhibitor. Furthermore, a network of hydrogen bonds was requested linking the second aspartate with the nitrogen atoms of the inhibitor, establishing a possible protonation pathway to be studied later. A detailed description of these hydrogen bond networks is given individually for each system below.

3.6.1 Pyrrolidine-protonated Inhibitor

For the unbound inhibitor model carrying a single proton, the pK_a values of the two tertiary amines indicate the hydrogen to be located at the pyrrolidine N-atom. Nevertheless, the formation of a covalent bond between the enzyme and the inhibitor according to the proposed mechanism leads to a ring opening of the aziridine moiety. Thus the basicity of the aziridine nitrogen center upon binding would drastically increase, which could only be compensated for by a stabilization via a proton shift. The only available proton source within the current system is the pyrrolidine function. Despite their close proximity, a direct transfer of the proton between the two nitrogen atoms is quite unlikely. Therefore, the molecular dynamics simulation trajectory was searched for geometries, in which one or two molecules are arranged such as to accept the H34 hydrogen atoms from the pyrrolidine nitrogen and to donate one of their owns to the aziridine N-atom. A suitable arrangement was frequently encountered during the simulation. However, this mechanism alone would require the inhibitor to already be protonated before the actual bond formation process can start. Hence, the geometries in question were further screened for a possible pathway connecting the H34 atom to the Asp25' residue. With a protonation pathway between these two residues, a protonation of the assumed neutral inhibitor molecule by the Asp25' residue could be computed in a reverse manner by shifting H34 away from the pyrrolidine N-atom. Due to the sterical restraints of the bicyclic core of the inhibitor, a direct transfer towards the Asp25' residue could not be expected, hence another water molecule was required bridging the H34 hydrogen atom an the carboxylate function. While the number of suitable geometries was largely reduced due to this additional requirement, suitable geometries were still available throughout the complete trajectory. Among the possible candidates, I have selected frame number 792 for the further calculations. Its geometrical arrangement is sketched in Figure 3.16, which

depicts an almost strain free proton transfer path between the two nitrogen atoms of the inhibitor using two intermediate water molecules and a second hydrogen bond path of equal length linking the N-atoms with the carboxylate of Asp25'.

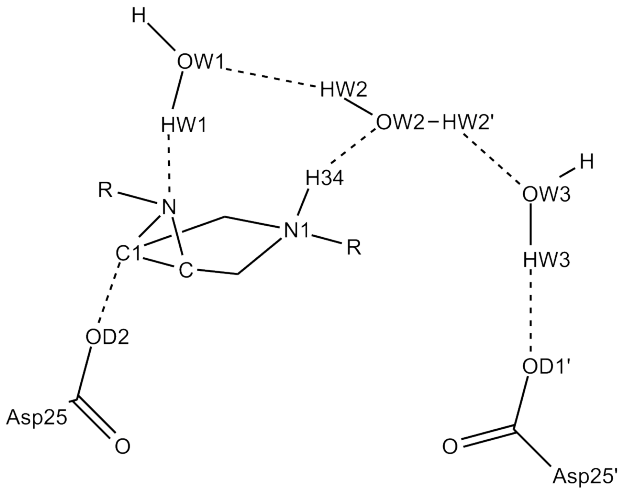


Figure 3.16: Chemical structure of the QM-subsystem used in the reaction path calculations on the pyrrolidine protonated inhibitor model. Water molecules two and three serve as proton wire between Asp25' and the pyrrolidine while molecule one can facilitate the shift of a proton onto the aziridine nitrogen center.

The geometry taken from the molecular dynamics simulations was prepared for the QM/MM calculations by firstly truncating the water shell to a sphere of only 15Å radius around the inhibitor core in order to reduce the size of the system and thus the computational effort needed. The actual QM/MM calculations were then performed on an even smaller region of 12Å around the inhibitor core, using the surrounding atoms as boundaries with their respective coordinates being unchanged in the calculations. The quantum mechanically treated subsystem was chosen to comprise the side chains of the two catalytic residues, truncated between their C_α and C_β atoms, the three water molecules depicted in Figure 3.16 as well as a largely truncated inhibitor model. Thus the inhibitor is represented by its central 3,6-diazabicyclo[3.1.0]hexane scaffold with the substituents truncated to a methyl group at the pyrrolidine and an ethyl group at the aziridine N-atom. To approximate the reaction path of the inhibition reaction, the reaction coordinate was approximated by two interatomic distances, that allow to follow the path of the reaction in a step-wise manner. The first coordinate that was chosen is the distance between the nucleophilic oxygen atom of the Asp25

residue and the C1 atom of the inhibitor. This coordinate is used as a direct measure for the formation of the covalent bond between the enzyme and the inhibitor. The second coordinate was used to model the proton transfer from the pyrrolidine to the aziridine nitrogen atom. Thus the interatomic distance between the N-atom of the aziridine and the HW1 hydrogen atom was used. Technical difficulties were to be expected in describing the proton transfer along the hydrogen bond network between the two nitrogen centers by a single interatomic distance. Previous investigations have shown that if only one of the numerous distances along this path is modified, the result are often the emergence of oxonium or hydroxyl ions that cause the energies of intermediates and transition states to be largely overestimated. The modification of a single distance does not account for the usually almost simultaneously occurring proton transfer between each pair of participating hydrogen bond partners, each transfer having its own transition state and activation barrier. To resolve these problems, still a single distance was selected as coordinate for the potential energy surface and being constrained during the individual calculations, nevertheless, the modifications of the molecular structure between each point along the surface and the adjacent one involved additional modifications. The major component of the proton transfer is the protonation of the aziridine nitrogen atom, described by the distance between the N-atom and the HW1 hydrogen atom. Nevertheless, when decreasing the distance between these atoms, also the distances between OW1 and HW2 or between OW2 and H34 were reduced. However, no constraints have been imposed on them during the QM/MM calculations. The result of this procedure is that the positions of HW2 and H34 are biases towards geometries in which the proton transfer is either completely on the side of the pyrrolidine or the aziridine nitrogen. Yet the exact positions of the hydrogen atoms are unconstrained, such that they can freely adopt minimizing the potential energy. Computations carried out on similar mechanisms have shown, that this procedure yields plausible reaction paths and energies without the emergence of unrealistic structures such as free oxonium of hydroxyl ions.

The potential energy surface given in Figure 3.17 exhibits three minima, two of which are well defined while the third one is not. This region corresponds to the non-covalent complex of the enzyme with 1020 while the proton still resides on the pyrrolidine nitrogen center. The geometrical fluctuations within this minimum can be explained by a rather large flexibility of the water molecules two and three shown in Figure 3.16 and a third molecule in their vicinity. Due to the differences in the energetical description especially of the hydrogen bonds between the water molecules, the distances in the geometry of the initial structure do not correspond to a minimum within the QM/MM description used to generate the potential energy surface. Hence, to minimize the total energy within the framework of QM/MM, the water molecules rearrange such as to optimize any interactions within the system

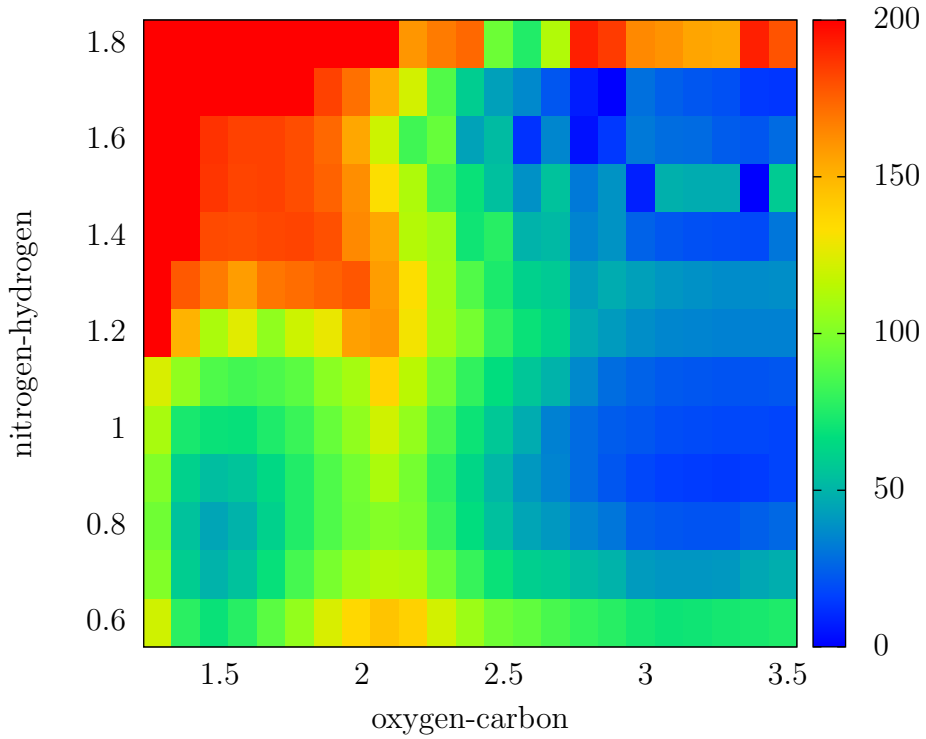


Figure 3.17: Top-down representation of the potential energy surface for the pyrrolidine-protonated inhibitor within the deprotonated active site of HIV-1 Protease (minimum on the top right). Additional minima depict the proton transfer to the aziridine nitrogen atom (bottom right) and the formation of the covalent complex (bottom left). Coordinates are given in Å, energies in $\frac{kJ}{mol}$ with respect to the lowest structure (3.4Å/1.5Å).

which requires changes in the coordinates of some atoms in the range of up to 1.0Å. These geometrical deviations are also facilitated by the choice of the coordinates, which leave the water molecules, except for the HW1 atom, almost unconstrained. Nevertheless, this was not considered to be a problem for the ongoing investigations, as following the reaction coordinate, the calculations converged to clearly defined minima for the non-covalent complex of the aziridine protonated inhibitor (lower right of Figure 3.17) and the desired covalent complex (lower left). To give an estimate for the reaction energy and activation barriers, the lowest energy obtained was taken as a reference ($\pm 0 \frac{kJ}{mol}$). This global minimum was found at the coordinates 3.4Å/1.5Å. As expected from the very first QM model calculations, the beginning of the reaction mechanism can be simplified to the proton transfer onto the aziridine nitrogen center. The corresponding minimum at 3.3Å/0.9Å is well defined and has a relative energy of about $+20 \frac{kJ}{mol}$ above the initial one. This is in good agreement with the differences in the pK_a values of the two heterocycles indicating a clear preference for the protonation of the pyrrolidine ring. The activation barrier for this proton transfer was estimated to be about $+34 \frac{kJ}{mol}$ which would agree well with proton transfer reactions being relatively fast at ambient temperature. Once the aziridine is protonated, the bond formation between the enzyme and the inhibitor can take place. Conversely to the previous step, the remainder of the mechanism is solely determined by the first coordinate while the changes in the second one are almost negligible. The structure of the covalent complex was found at the coordinates 1.5Å/0.9Å, exhibiting a relative energy of $+54 \frac{kJ}{mol}$ with respect to the global minimum. The transition state for the bond formation was found at 2.1Å/0.8Å with and activation barrier of $+66 \frac{kJ}{mol}$ with respect to the preceding minimum. With the overall reaction energy of roughly $+55 \frac{kJ}{mol}$, the formation of a covalent bond starting from the singly protonated inhibitor is energetically strongly disfavored and will not take place. This results however does not necessarily imply that the proposed molecule can not be used as inhibitor as described below.

3.6.2 Doubly protonated Inhibitor

The second system, for which the molecular dynamics simulations yielded promising geometries for the formation of a covalent enzyme-inhibitor complex comprises two protons, each one of the located at either of the two nitrogen atoms of the 3,6-diazabicyclo[3.1.0]hexane scaffold. Selecting the most suitable geometries was guided by similar requirements as those previously described. Again, the major criterion for a suitable frame was a short distance between the nucleophile and either of the two bridgehead carbon atoms of 1020. Yet due to the presence of a second hydrogen atom, a different structure of the hydrogen bond network was desired. With both tertiary

amines being protonated, an intramolecular proton transfer was neither required nor possible. However, modeling such a transfer between the inhibitor and the Asp25' residue was still highly desired. Hence the most promising geometries exhibit an interaction pattern similar to the one sketched in Figure 3.18 with the 196th frame being the one used further on.

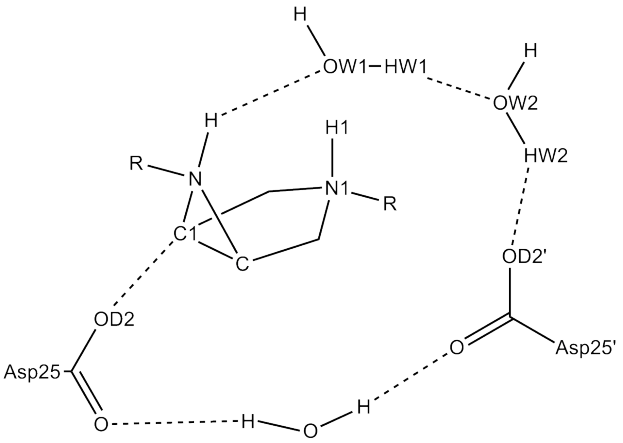


Figure 3.18: Chemical structure of the QM-subsystem used in the reaction path calculations on the doubly protonated inhibitor model. The water molecules one and two are used to model the proton transfer reaction between the Asp25' residue as the aziridine nitrogen center preceding the bond formation and ring opening.

Again, the sterical demands of the bicyclic core of the inhibitor prevent a direct hydrogen bond between the aspartate and either of the protonated amines, thus the working system again comprises a proton wire of two intermediate water molecules. Whereas water molecule two acts as proton donor towards Asp25', molecule one behaves as acceptor for both of the NH bonds of the inhibitor. This network of hydrogen bonds is ideally suited to model the proton transfer between the Asp25' and either of the N-atoms of the inhibitor. However, as only the one with the aziridine nitrogen center was investigated within this work, only the corresponding hydrogen bond is drawn in the graphic. For the computations on the potential energy surface, the leading coordinate to describe the bond formation was again chosen to be the distance between the nucleophilic oxygen of Asp25 and the C1-atom of the inhibitor. To model the proton transfer, the second coordinate was selected to be the distance between the aziridine nitrogen atom and its proton. Whereas exclusively these two interatomic distance had been used for the actual calculations, two more distances have been used while modifying the geometries between adjacent points of the potential energy surface. The proton transfer between the nitrogen atom and the Asp25' residue was again

biases to either side of the reaction by adjusting the distance between OW1 and HW1 as well as that between OW2 and HW2 simultaneously and to the same degree as that of the NH function. Nevertheless, these distances were left to freely relax during each individual calculation.

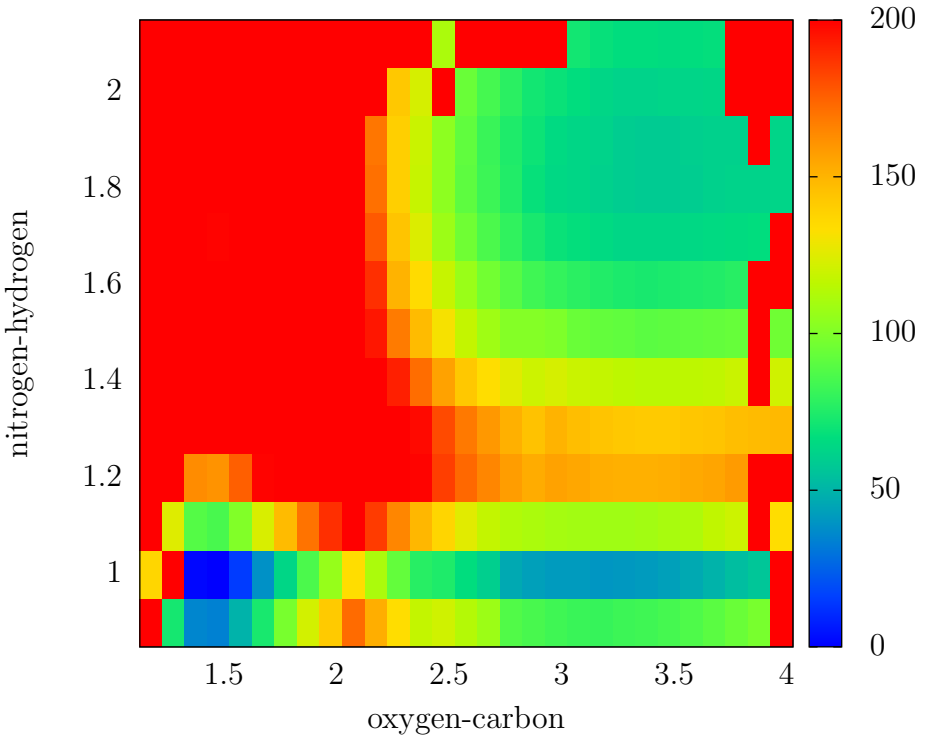


Figure 3.19: Top-down view on the potential energy surface computed for the doubly protonated inhibitor and deprotonated active site (bottom right). The covalent complex is shown as minimum on the lower left while the minimum with the proton shifted back to the Asp25' residue is depicted in the upper right. Coordinates are given in Å, relative energies in $\frac{kJ}{mol}$ with respect to the covalent complex.

The resulting potential energy surface is given in Figure 3.19, which exhibits three clearly distinct and also well defined minima. The minimum on the bottom right corresponds to the simulated system with the deprotonated catalytic dyad and the doubly protonated inhibitor. Following the reaction along the decrease of the oxygen-carbon distance, one reaches the global minimum on the left side of the surface representing the covalent adduct of the inhibitor and the enzyme. The shallow third minimum, depicted on the top right side of the graphic is the result of moving the proton from the inhibitor towards the Asp25' residue. Hence this structure could also be imagined to

be the starting point of the reaction if the non-covalent complex would have been formed from the enzyme in its monoprotonated form and the inhibitor being solely protonated on its pyrrolidine N-atom.

Not surprisingly, the proton transfer between enzyme and inhibitor can be defined via the change in the N-H distance only while that between the Asp25 residue and the inhibitor is hardly affected. The energetics of this reaction are also in good agreement with the expectations of the protonated inhibitor being more favorable than a carboxylic acid function in vicinity to a tertiary amine. The calculations estimate the protonated inhibitor to be $-18\frac{kJ}{mol}$ more stable. The proton transfer can be expected to proceed rather easily, as the activation barrier amounts to only $+17\frac{kJ}{mol}$. This is in contrast to the value published earlier.[161] The difference originates from the estimated position of the transition states, which are 0.1Å apart from each other considering the oxygen-carbon distance. Nevertheless, the slope of the potential energy surface along this direction is extremely steep, leading to significant deviations in the relative energies. Nevertheless this does not affect the interpretation of the overall inhibition reaction as, which will be shown later, the rate determining transition state is energetically significantly above both estimated values. From the minimum structure of the doubly protonated inhibitor, the formation of the covalent enzyme-inhibitor complex proceeds via a decrease of the oxygen-carbon distance from 3.3Å to only 1.5Å once the bond is formed. The activation energy for this reaction is $+46\frac{kJ}{mol}$, allowing the reaction to happen at a reasonably high rate at ambient temperature. In contrast to the previously studied system, the bond formation reaction is exothermic, although the reaction energy of only $-51\frac{kJ}{mol}$ implies a reversible rather than the expected irreversible process. Although a decreased reaction energy in comparison to the ones determine for EPNP or AZNP imply lower pK_i values, this also leads to a reduction of possible side effects, one of the most severe concerns against the development of covalent drugs.

3.7 Conclusions

The protocol developed for the rational design of covalent inhibitors combines classical and quantum mechanical methods to describe the reactivity between the model compounds and the target enzyme with well established docking procedures which are used to optimize the binding affinity with the target protein. Starting from Xray crystal or NMR structures of complexes with proteins and covalently bound test molecules, our methods allow to derive a working model for the reactive species. These can then be used as lead structures to optimize binding affinities or selectivities by modifications of their chemical structure. Furthermore, tuning of the reactivity against the target enzyme is possible via modifications of the functional group that co-

valently reacts with the protein. The application of the developed protocol on potential covalent inhibitors against HIV-1 Protease yielded promising model compounds which are currently being synthesized and tested in vitro against their target enzyme.

4 Influences on the Electron Density of Molecules

The interactions of molecules with their immediate surroundings are based on the same set of physical laws, no matter what the exact nature of the environment might be. Therefore, comparable interactions can be found between two molecules in vacuum, solvated species and the surrounding liquid, molecules and their crystal environment or even between smaller chemical compounds and large bio molecules such as proteins. As the interaction pattern in chemical systems is always arranged such as to minimize the total energy of the system, comparable patterns are to be expected for the same molecule but changing the environment from e.g. a protein complex to the solvated species. In order to maximize stabilizing contribution, molecules are expected to adopt as far as possible to form and optimal interaction pattern with their environment, most apparent in the picture of the lock-and-key principle for describing the affinity towards target enzymes. Conversely, the environment might also adapt to the electronic structure of a small molecule observed within numerous studies on the solvation structure of ions or other small compounds. The idea of an optimal interaction pattern between molecules and their surrounding has led to the common practice to approximate the investigations on the interactions between enzymes and their substrate or inhibitors by using the corresponding crystal structure of the pure compound as a model.[16,17,172] The latter can usually be obtained far more easily than that of protein complexes and, due to fewer defects and disorder in the crystal, the resolution is generally far better than that obtained for the protein system.

A direct correlation between the interaction patterns of pure crystals of ligand molecules and protein complexes however is by no means obvious. Furthermore, only a very limited number of investigations has been carried out so far, that compare the situation of molecules in these two environments with each other.[18] The obtained results yield ambiguous results. The study reports a good resemblance between the crystal structure of a non-covalent inhibitor,[173] Tranexamic acid or AMCHA, with the structure determined inside its target enzyme (pdb ID 1ceb[174]). With the molecule residing in a zwitterionic form, providing a carboxylate and an ammoniumion, the are several strong hydrogen bonds between adjacent molecules within the pure crystal. Interestingly, an identical set of hydrogen bonds is formed within the

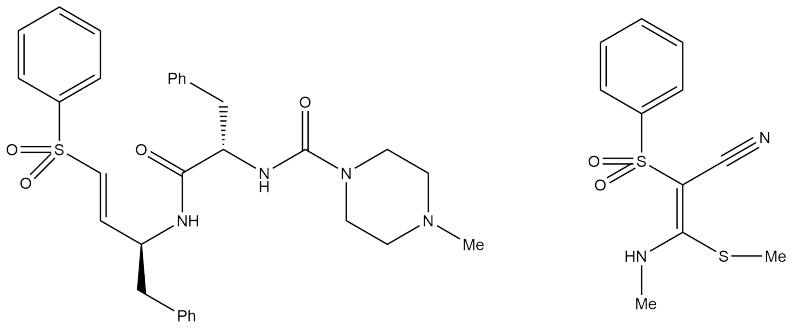
recombinant kringle 1 domain of human plasminogen, although the bonding partners are exchanged for the amino acid residues of the protein. The second investigated model was the non-covalent complex between Cathepsin B and loxistatin acid (E64c), an epoxide-based, covalently acting inhibitor.[175] The bonding situation was compared to the crystal structure of loxistatin (E64d),[176] the ester derivative of E64c. For both molecules, the hydrogen bonding network with the peptidomimetic majority of the molecule was again found to be very similar. A more detailed look at the functional group that is responsible for the covalent interaction with the target however revealed significant differences in the electronic structure of the epoxide function which is caused by the unique binding situation within the protein. The interactions within the active site require a geometry, in which the attacking nucleophilic amino acid residue is extremely close to the epoxide and one of its carbon-oxygen bonds is already considerably weakened.

With the exception of the unique binding situation of the covalent inhibitor, the crystal structures of the pure compounds were concluded to be suitable good approximations to the binding situation within the protein complexes. If the differences found for the non-covalent enzyme inhibitor complex between E64c and Cathepsin are indeed a special case or possibly a feature of any covalent inhibitors was not clarified. Hence, the current investigations are carried out for another class of molecules that are used as covalent inhibitors addressing cystein proteases.

4.1 The Model Compound

There is a multitude of potential functional groups, that could be used as central scaffold in the rational design of novel covalently acting inhibitors. These groups are usually electrophilic allowing to react with nucleophilic active site residues such as cysteines, serines or aspartates. One functional group that has emerged as promising candidate for further developments are vinylsulfones. Experimental studies have shown that vinylsulfones can efficiently inhibit cystein proteases such as cathepsin, rhodesain or falcipain, while side reactions with proteases of other classes are only minimal.[177,178] The K11777 molecule in Figure4.1 is one of the most potent and best studied compounds of this new class of inhibitors. Furthermore, crystal structures of K11777 with three representative enzymes of cysteine proteases have been published recently[179], which provide a solid basis for a rational and structure based design process.

Whereas the binding affinity of K11777, like for other inhibitors as well, originates from the interactions of its recognition unit with the corresponding binding pockets, it is the vinylsulfone scaffold which is in the focus of current investigations.[180] The immediate substitution pattern of this group can be



K11777 TS793

Figure 4.1: Chemical structure of K11777,[178] an inhibitor against Cruzain, Rhodesain and Falcipain[179] as well as that of TS793, a smaller model compound used in theoretical investigations on the reactivity of vinylsulfones against cystein proteases.[180]

used for tuning the reactivity against the catalytic cysteine. Hence, investigations on the inhibition reaction of Rhodesain by K11777 were accompanied by a screening of several other small vinylsulfone-based compounds, one of them being TS793 also shown in Figure 4.1. These investigations comprise theoretical computations on the inhibition reactions as well as the synthesis and crystal structure determination of each of these compounds. Therefore, the TS793 is also ideally suited for the investigations on the electron density and its influences arising from various surroundings.

4.1.1 TabuSearch using Charmm

In order to estimate the environmental influences on the TS793 molecule, it is necessary to know the properties of the isolate molecule in the first place. Therefor, the molecule was subjected to a scan over its conformational space to determine all possible, thermodynamic minima and their energies. The latter can then be used to estimate the population of each of the minima assuming a Boltzmann distribution.[181]

The sampling of the phase space of the molecule was carried out using a recently developed program, CAST.[46] The implemented algorithms in combination with computationally inexpensive methods allow for rapid yet thorough scanning of the conformational space. Hence, the first scan was performed employing the Charmm force field[27] using the topological definition and parameter set generated using the Charmm-GUI interface.[163] As the atomic charge assigned during this process are known to be of only modest quality,[182,183] they have been replaced by AIM charges calculated

for the isolated molecule. The AIM charges have been preferred over the Mulliken charges, although the latter are used in the Charmm general Force Field procedure.[165] Nevertheless, besides the large basis set dependence of the Mulliken charges, the AIM charges are also less sensitive to the exact conformation of the molecule. Therefore, using the AIM charges is expected to provide are less biased basis for scanning the conformational space. The conformational space of TS793 was expected to exhibit only little diversity, as the molecule does only posses six rotable bonds, which could potentially lead to new minima. Furthermore, rotations around three of these bonds are not expected to add to the number of physically relevant minima, as rotating the R_1 substituent by $\pm 180°$ or rotating either R_2 or R_3 by $\pm 120°$ will lead to chemically identical structures due to the two- and threefold symmetry of the terminal phenyl- and methyl-groups. Thus, a limit of 1000 iterations was deemed sufficient for the computations.

The scan performed with CAST terminated well within the allowed 1000 iterations, indicating that all physically relevant conformations have been sampled. The total number of minimum geometries obtained amounts to 805, thus the visual inspection was supplemented by a variety of statistical measures. The central vinyl scaffold, comprising the S_1, S_2, N_6, C_7, C_8 and C_9 atoms, was used to mutually align each pair of geometries along the trajectory for measuring similarities via the standard deviation in the atomic position of the nuclei. The twofold measurement depicted in Figure 4.2 was obtained by measuring the RMSD values for the above mentioned substructure as well as for the entire molecule.

The dominant feature in Figure 4.2 is the strict partitioning in either extremely low (blue) or high (red) values within the lower half of the graph. A correlation of these values with the molecular geometries revealed, that the obtained minima do not only cover the conformational space of the original Z-conformer, in which the S1 and S2 atoms are located opposite to each other, but also the E-conformer, resulting from a rotation of the C7=C8 bond. Low (blue) RMSD values thus represent comparisons of Z- with Z-conformers (or E- with E-conformers), which can readily be superimposed with each other. In contrast to that, comparisons between Z- and E- conformers result in large (red) RMSD values, as a superposition of the two is not possible. The Z-conformers are physically irrelevant, as the energy barrier of the rotation around the C7=C8 double bond is far too large to be surmounted. Thus, all Z-conformers have been removed to facilitate further analysis. Nevertheless, having them found by CAST is again a strong sign, that the sampling of the conformational space was complete.

Another feature of Figure 4.2 is the checkerboard-like structure, most prominently within the clusters located right above the diagonal but also faintly visible in the entire graph. Each cluster exhibits lower (blueish) and mediocre (green to yellow) RMSD values, indicating two distinct groups of

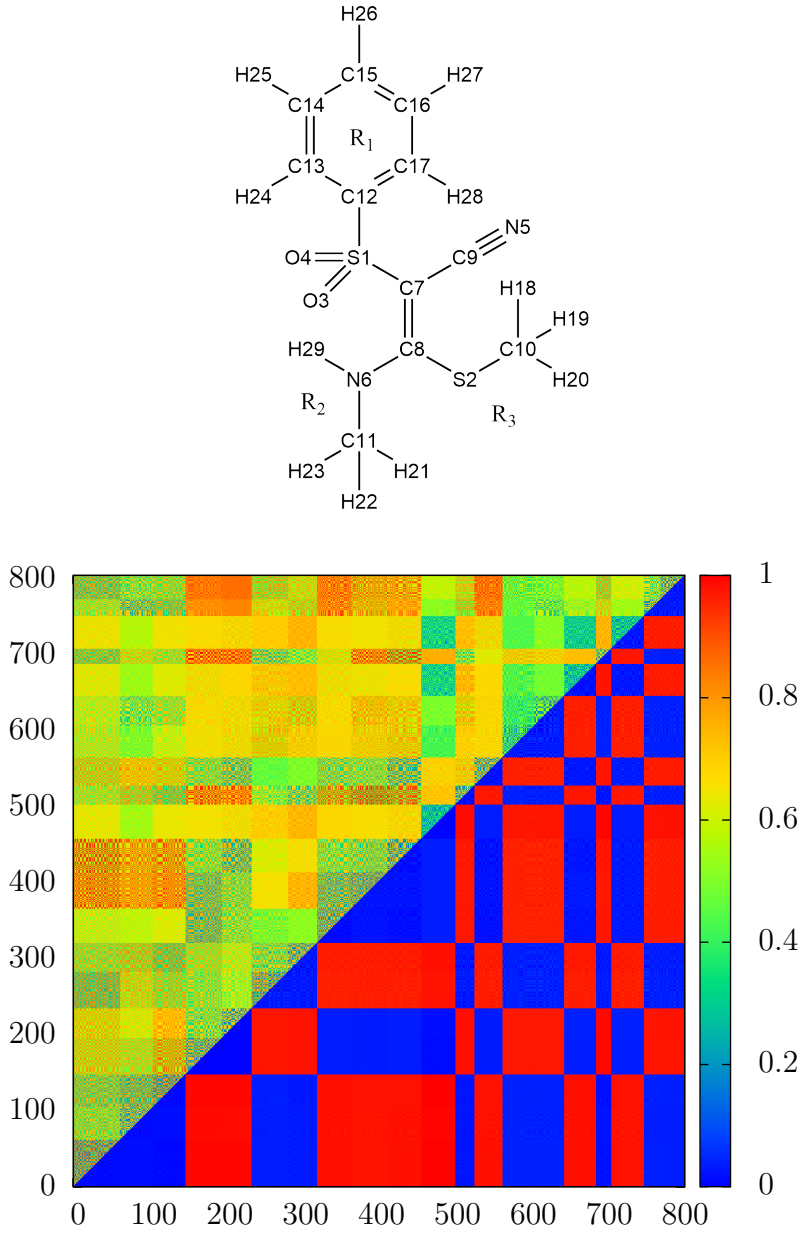


Figure 4.2: Top: chemical structure of TS793 with atomic labels; Bottom: Two-dimensional representation of RMSD values measured along the trajectory of 805 geometries obtained using CAST and the Charmm force field. The alignment (bottom right) was performed based on the S1, S2, N6, C7, C8 and C9 atoms while the RMSD measurement (top left) was performed for all atoms. Due to the largely different ranges, only relative RMSD values are given, normalized to the largest one within each half.

geometries. Whereas the geometries within one of these groups are almost identical, there is a significant difference between those between the two groups. A visual inspection of each cluster revealed, that this bifurcation is caused by a specific twofold symmetry of TS793. The central scaffold comprising the atoms S1, S2, N6, C7, C8 and C9 is almost planar and thus, with the substituents R_1 to R_3 being located "above" or "below" them, acts as a symmetry plane. The result is that, even though the physical properties such as the total energy are the same, every cluster contains a specific geometry as well as its mirror image. Hence, to further simplify the analysis, those geometries with a negative dihedral angle C8-C7-S1-C12 have been inverted, such that the R_1 substituent is always located "below" the vinyl scaffold.

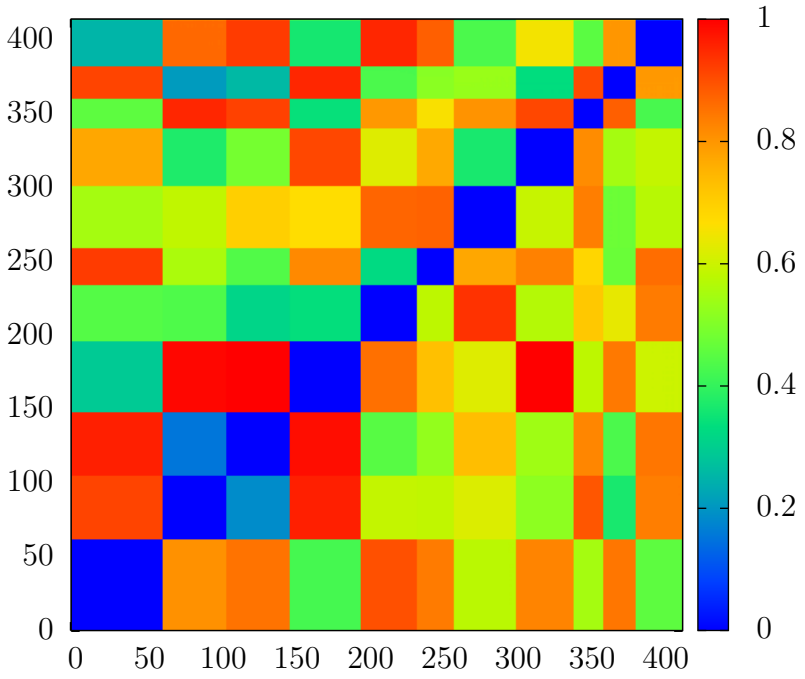


Figure 4.3: Two-dimensional representation of RMSD values measured among the remaining 414 geometries of the E-conformers. The alignment (bottom right) was performed based on the S1, S2, N6, C7, C8, C9, C10, C11 and C12 atoms. To avoid biases due to rotational symmetry of the R_1 to R_3 substituents, the measurements of RMSD values were carried out only for symmetrically unique atoms. For better visibility, values within each half have been normalized to the largest value.

Following the removal of the Z-conformers and the inversion of some of the

geometries, another measurement of RMSD values was performed. Again, the central vinyl scaffold was taken as a reference while the measurement was performed on the entire molecule. The graphical representation in Figure 4.3, visualizing the measured RMSD values, provides a clear separation of the trajectory in twelve clusters, indicated by the blue triangles (RMSD values of 0.0 Å) below the diagonal. The last of the clusters is hardly visible, as it contains only a single geometry. Despite the perfect agreement in the atomic positions of the vinyl scaffold, the RMSD values measured for the entire molecule again exhibit a variety of values. Unlike before, there are more than two distinct values within each of the clusters. These differences originate from rotations of the terminal phenyl- and methyl-groups. Whereas rotations of $\pm 180°$ for the former and $\pm 120°$ of the later do not lead to physically different minima, the sequential indexing of the atoms causes CAST and index based measurement to identify them as separate geometries. Yet a visual inspection revealed, that despite the enumeration, all geometries within each cluster could be perfectly superimposed with each other. Thus using a single representative geometry for each cluster, the number of geometries could finally be reduced to twelve. In order to validate and correlate the geometries, a final two-dimensional RMSD measurement was performed. The alignment was again performed on the vinyl scaffold. The actual measurement was carried out on all symmetrically unique atoms, namely S1, S2, N5, N6, C7, C8, C9, C10, C11, C12, C15, H26 and H29. The graphical representation of the obtained values is given in Figure 4.4.

The energy range for these 12 geometries computed with the Charmm force field ranges to about $50\frac{kJ}{mol}$ above the global minimum ($0\frac{kJ}{mol}$), the energetically lowest one. A graphical representation of the relative energies is given in Figure 4.5. In addition to the relative energies, also the interatomic distance between the S1 an H29 atoms is shown. It can be used as a measure for the intramolecular hydrogen bond, which is formed between H29 and one of the two oxygen atoms of the sulfone (O3 or O4) in the energetically favored geometries 1 to 3. Concerning the relative energies, the hydrogen bond has an absolutely dominating effect, energetically separating those conformers, where it is present by at least $25\frac{kJ}{mol}$ from those, which lack this stabilization. Among the first three geometries, which exhibit this hydrogen bond, it also preorganizes the orientation of the substituents, such that the R_1 and R_2 substituents are almost identical between them. The only significant difference can be found in the orientation of the R_3 substituent, which is "below" the molecular plane in geometry 1 and "above" in the other two. Nevertheless, its influence on the energy is almost negligible.

Assuming a Boltzmann-like distribution between the possible conformers, those lacking the hydrogen bond with relative energies of about $30\frac{kJ}{mol}$ or more are not expected to be significantly populated (less than 5 ppm at 293 K). Furthermore, due to the potentially non-optimal force field parameters, the

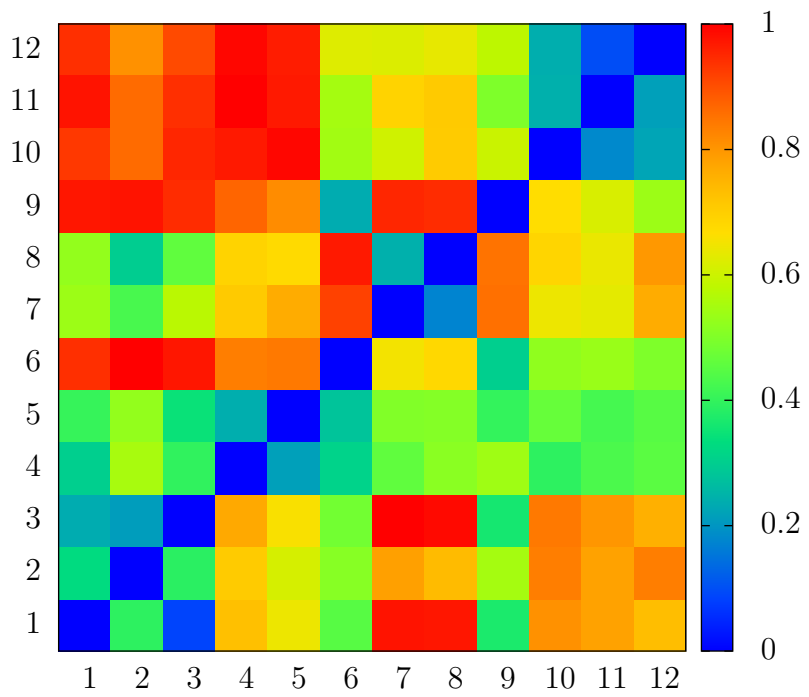


Figure 4.4: Two-dimensional representation of the relative RMSD values measured for the twelve final geometries obtained by CAST. The alignment was performed using the S1, S2, N6, C7, C8 and C9 atoms. The additional measurement was performed on all symmetrically unique atoms. For better visibility, all values in the triangles above and below the diagonal have been normalized to the largest value, respectively.

energetically elevated geometries exhibit a significant pyramidalization at the N6 center. Hence, a refinement of all twelve geometries was performed using the Turbomole program.[42] The energies have been evaluated using density functional theory employing the B3-LYP functional.[36,38,184] Two series have been carried out, one computing the isolated molecule and the other one approximating liquid, aqueous surrounding by means of a polarizable continuum method, COSMO.[185]

Not surprisingly, the refinement of the geometries obtained with the Charmm force field by means of quantum mechanical calculations leads to considerable changes in each of them. Nevertheless, as depicted in Figure 4.5, the overall correspondence between geometry and total energy is comparable within each of the three trajectories. Significant changes however occurred for the frames 6 and 9, which relaxed to much more favorable geometries both within the vacuum and solution models. The large stabilization stems from a rearrangement, which enables the formation of the above mentioned hydrogen bond between the sulfone and H29, represented by the interatomic distance between the hydrogen and the corresponding sulfur atom S1 (Figure 4.5).

4.1.2 TabuSearch with Semi empirical Methods

Whereas the first scan using the Charmm force field was computationally very efficient, there remain uncertainties due to the assigned force field parameters. Hence, a second scan with CAST was started, employing the MOPAC program for the energy evaluations based on the semi empirical PM7 method.[45,186,187,188,189,190] Besides the exchange of the energy functional, the parameters for the CAST program have been left identical to those used in the first scan. Employing the PM7 method, a total of 56 minimum geometries was found out of which six geometries could be removed right from the start. As already experienced in the previous search, these geometries correspond to the Z-conformer of TS793 which is not in the focus of the current investigations. The remaining 50 minimum structures exhibit energies in a range of about 40 $\frac{kJ}{mol}$ as given Figure 4.6. In order to deliver results comparable to those of the preceding run, each geometry was further refined using Turbomole and DFT with B3LYP/TZVP as before. Again two series were conducted computing the molecule in vacuum or approximating aqueous solution using the COSMO with its standard parameters and an ϵ-value of 78.39 in consistency with previous calculations. The corresponding energies are also given in Figure 4.6.

Obviously, the first four minimum geometries are by far more favorable than the remaining ones throughout all methods used. The low energies originate from the intramolecular hydrogen bond, which can be found only for these four geometries. Due to this hydrogen bond, the methylamine sub-

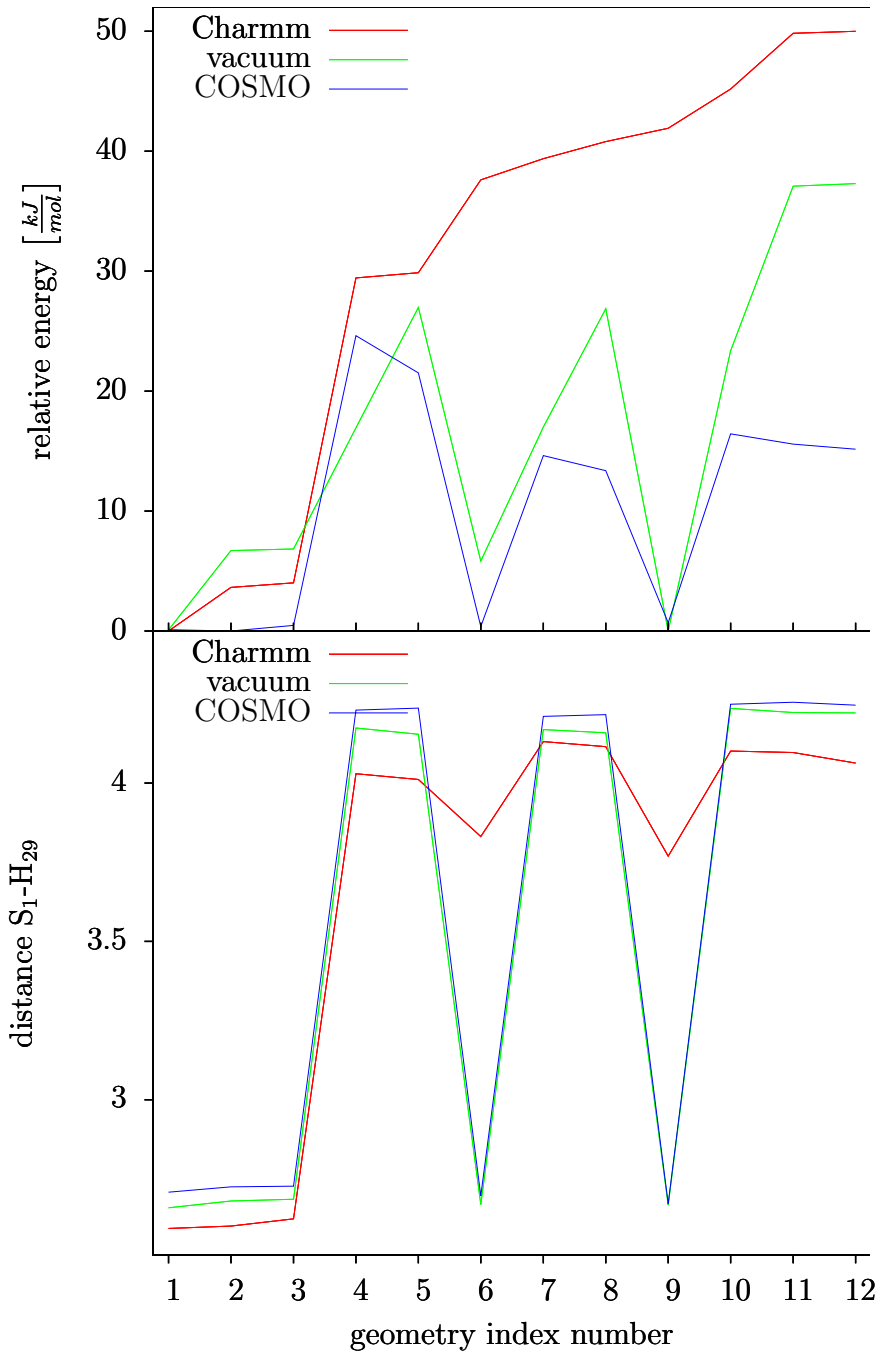


Figure 4.5: Top: relative energies of the 12 unique minimum geometries ranked by the energies obtained with Charmm; Bottom: iteratomic distance between the S_1 and H_{29} atom indicating the presence of absence of the intramolecular hydrogen bond. "vacuum" denotes a refinement with B3-LYP/TZVP while an additional solvent model was used for "COSMO".

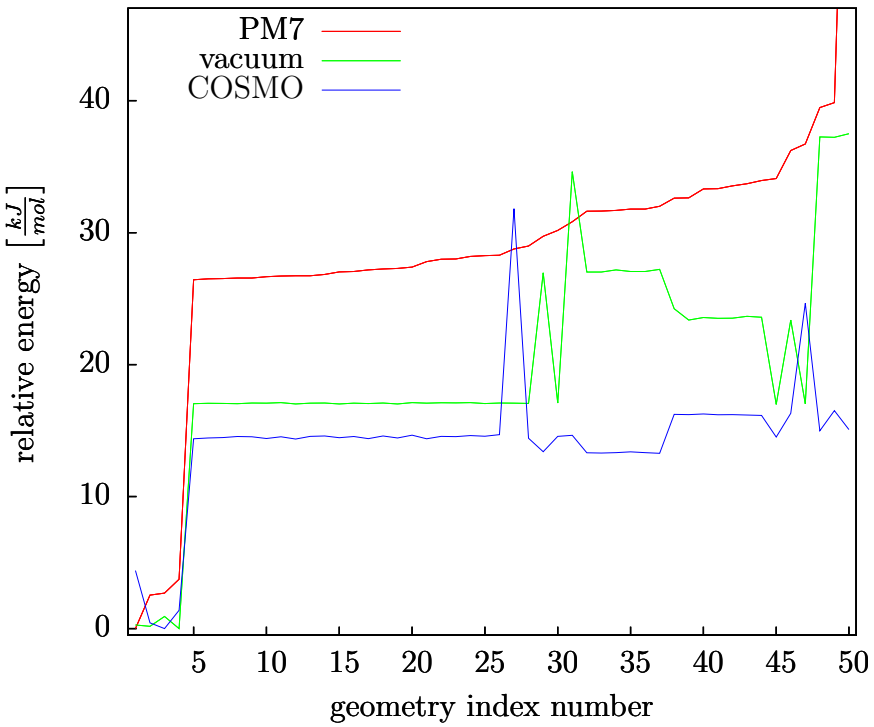


Figure 4.6: Minimum Energies of all 50 E-conformer geometries obtained from the TabuSearch run using MOPAC and the PM7 method as well as relative energies computed for the geometries refined in vacuum (pure DFT) and solution (additionally with COSMO).

stituent is oriented almost coplanar with the vinyl scaffold. Differences can be found in the relative orientation of the phenyl ring and the thiomethyl substituent which are in anti-conformation for the first but in syn-conformation in the next three minima. Although having an almost negligible effect on the overall energy, this causes the first minimum to be the global one for the PM7 method whereas both DFT approaches find the energetically lowest geometry among the almost identical geometries two to four. However, both minimum geometries correspond nicely with the two most favorable ones determined from the TabuSearch on the force field level. The same holds true for the energetically elevated geometries, which all correspond to geometries, which are missing the intramolecular hydrogen bond and thus its stabilizing effect. A cluster analysis of the obtained geometries showed, that using either Charmm or PM7 as energy functional during the TabuSearch run within CAST resulted in a complete scan of the conformational space of TS793 with a good correspondence between the obtained geometries. A further refinement of these geometries using DFT methods results in virtually identical geometries.

4.2 The Model Systems

In line with previous investigations,[18] the current study comprises theoretical calculations on the TS793 molecule in the non-covalent enzyme-inhibitor complex as well as within the crystal of the pure compound. The crystal model is based on the experimentally determined crystal structure of TS793 without major modifications. The protein complexes have been derived from the covalent complex of K11777 and are the final step of computing the bond formation back to the non-covalent complex. Additional calculations were carried out for the isolated molecule in vacuum as well as for the solvated species, approximating liquid aqueous surrounding by the use of the COSMO.[185]

4.2.1 Estimation of Reaction Energies

The TS793 molecule was selected as model compound for investigations on possible inhibition reactions with Rhodesain as a series of preliminary quantum mechanical calculations showed favorable thermodynamics among the potential reactions between the molecule and the active site cystein of the protein. The quantum mechanical model system contained the TS793 molecule and, to approximate the cystein and histidine residue, a methylthiolat and a proton.[180] The possible reactions between these compounds are sketched in Figure 4.7. They can be divided into two groups depending on the initial nucleophilic attack taking place either at the α- or β-carbon atom

of TS793. The emerging anion can then be protonated, leading to the addition products shown in the top and bottom center. Furthermore, any of the four substituents could be forced out as leaving group either directly from the intermediate anions (vinylic substitution) or as a result of an addition-elimination reaction from one of the addition products described before.

The substitution pattern of TS793 was guided by the idea of a covalent but reversible inhibitor which required the chemical reaction with the methylthiolate to be thermodynamically neutral. Hence, the TS793 also carries a thiomethyl substituent which, upon substitution, yields a reaction energy of exactly zero. Yet to assure that this reaction takes place, the reaction energies for the alternative reactions have to be higher. This was estimated using the quantum mechanical model with the obtained energies given in Table 4.1.[180]

reaction	added compound(s)	leaving group	reaction energy $\left[\frac{kJ}{mol}\right]$	
			minimal	alternative[180]
α-addition	MeS^-, H^+	none	+114	+80
substitution	$MeS^-(H^+)$	$PhSO_2^-/PhSO_2H$	+17	+17
substitution	$MeS^-(H^+)$	CN^-/HCN	+60	+67
β-addition	MeS^-, H^+	none	+57	+84
substitution	$MeS^-(H^+)$	$NMeH^-/NMeH_2$	+85	n.a.
substitution	$MeS^-(H^+)$	$MeS^-/MeSH$	± 0	

Table 4.1: Reaction energies estimates from the model system comprising TS793 and methylthiol. The reaction energy given in the last line refers to an isoenergetic reaction, thus it is by definition $\pm 0 \frac{kJ}{mol}$.

Although the energies determined by Schneider et al.[180] already depict the desired preference for the substitution of the thiomethyl substituent, the computations have been carried out only on selected conformations of the potential products. Therefore, another set of TabuSearch computations was started for each of the components appearing in Figure 4.7. The searches within the conformational space of each of the molecules were carried out identically to those described for the TS793 molecule itself. Due to the much smaller computational demands, only the Charmm force field was used in the actual TabuSearch calculation while all obtained minimum geometries have been refined using DFT employing the COSMO. Reaction energies were then estimated by using the minimum energies of reactants and products for each of the six reactions. These values are given in Table 4.1. Due to the use of additional conformers, some of the energies differ significantly from the ones previously calculated. Nevertheless, the general trend observed before is retained with any other reaction but the substitution of the thiomethyl

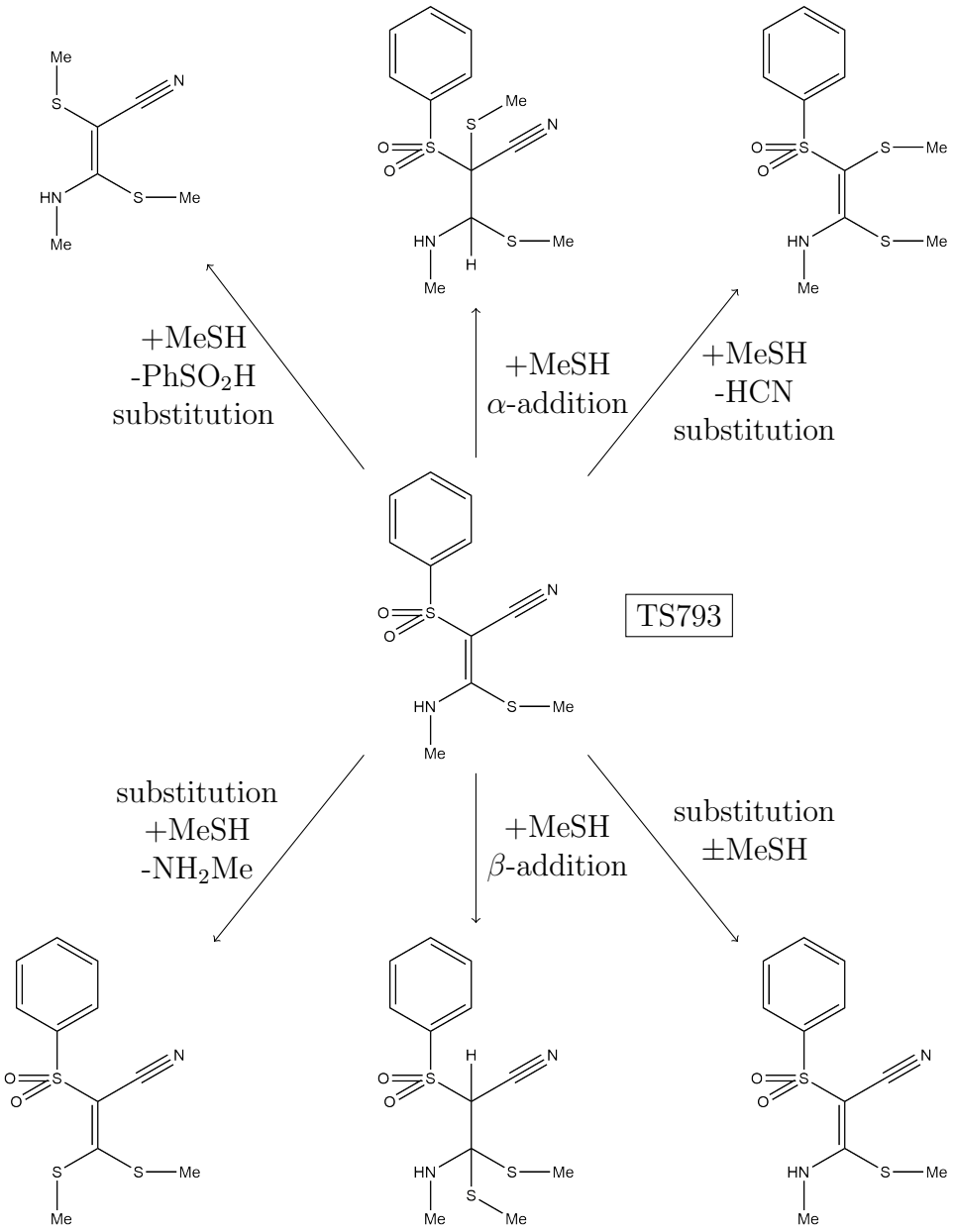


Figure 4.7: Chemical model reactions between TS793 and methylthiole comprising addition of methylthiolate in α- or β-position with respect to the sulfonyl-group and subsequent protonation (addition) or elimination of one of the former substituents (vinylic substitution). For substitution products, also an addition-elimination-mechanism is imaginable.

substituent leading to an endothermic reaction. Comparing the reaction energies of the α- and β-addition reaction, the almost twice as large value for the α-addition is in line with the crystal structures of K11777 in its enzyme-inhibitor complex, which is also the product of the more favorable β-addition. Although disfavored by the energy difference between reactants and products of the corresponding model system, the β-addition can still be expected to be an alternative reaction to the substitution which could further increase the residual time of the inhibitor. Hence, the complexes of TS793 with Rhodesain still represent an interesting model system to be studied with respect to the reaction mechanism as well as for more detailed investigations on the mutual interactions between the enzyme and inhibitor on an electronic level.

4.2.2 The Crystal Environment

A necessary prerequisite for the comparison different environmental influences was the existence of a suitable crystal structure of the pure compound. For the TS793 molecule, such a crystal structure has recently been determined and was used as a basis for the preparation of the corresponding theoretical model.[21] The crystal structure of TS793 has a monoclinic unit cell with only minimal deviations from a perfectly cuboid shape ($\beta \approx 99.943°$). The cell vectors have lengths of 7.9Å, 11.3Å and 14.1Å, and the unit cell contains four symmetry related copies of the TS793 molecule. Fortunately, no disorder has been found such that all atomic coordinates are unambiguously defined.

Within the crystal structure of TS793, the molecule can be found in a geometry closely resembling the global minimum structure found during the TabuSearch investigations on the isolated molecule. Therefore, almost no energy in required for the molecule to adopt a favorable geometry prior to the crystallization process. Furthermore, this geometry allows the formation of dimeric structures between two molecules of TS793. These dimers are held together by two strong hydrogen bonds between each pair of sulfone and methylamine substituents as depicted in Figure 4.8. An interesting feature of the hydrogen bond network is an almost equal length of the intramolecular (2.14Å) and intermolecular (2.26Å) hydrogen bonds. A more detailed analysis on them will be given further down.

Crystals are almost by definition periodic with respect to their molecular structure, thus using a computational model that exploits this periodicity seems an obvious choice. Nevertheless, the results obtained for the crystal model should then be compared with the other system which all lack a periodic symmetry. Hence, also for the crystal no periodic model was used. In line with the calculations used in the investigations of protein complexes, a representative fragment of the crystal structure was generated with a small

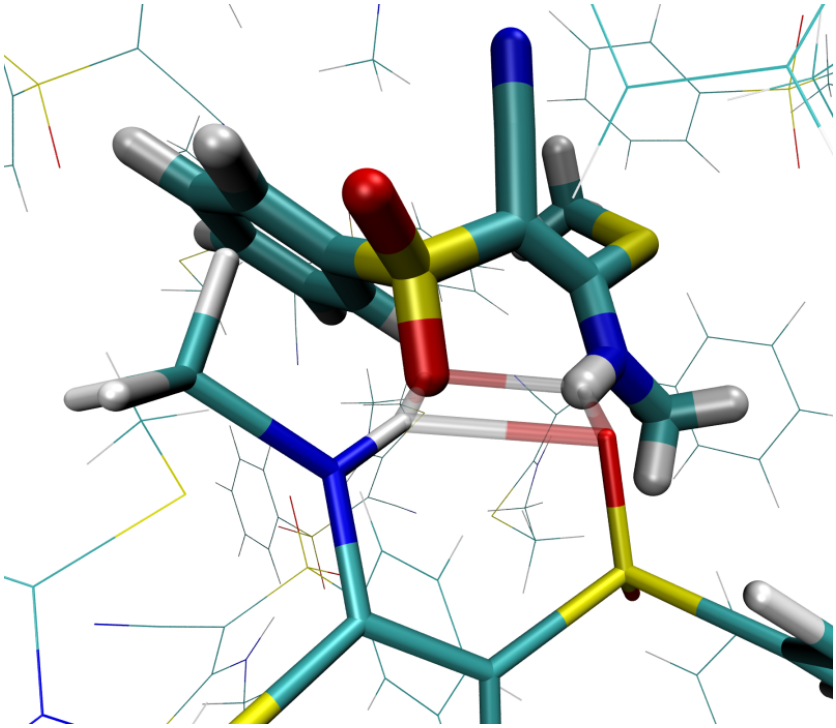


Figure 4.8: Sketch of the crystal structure of TS793. Dimers are formed via two strong hydrogen bonds between pairs of TS793, indicated by transparent bonds. Further interactions with surrounding molecules are relatively unspecific.

region comprising a single up to a few individual molecules being described by density functional theory while treating any of the surrounding molecules by means of classical mechanics only. In the case of TS793, building up the crystal was a straightforward procedure, as the crystal structure of that molecule only comprises only one chemical compound without further solvent molecules, ions or other additives. Hence, the crystal fragment was built using the new `Decifer` program, resulting in system compatible with the force field programs used. The parameter set for the TS793 molecules was identical to that used in the TabuSearch investigations before with the molecular topology and parameters being generated using Swissparam [164] and the atomic charges replaced with the AIM charges calculated separately. In order to avoid biases due to boundary effects, the crystal fragment was constructed from 1815 unit cells (15×11×11) resulting in a roughly cuboid structure of 115Å×115Å×145Å, containing 7260 molecules or 210,540 atoms.

4.2.3 The Protein Environment

As mentioned before the TS793 is derived from the K11777, thus also its enzyme-inhibitor complexes have been calculated from a modified structure with the crystal structure of K11777 and Rhodesain (pdb ID 2p7u [179]) being used as a basis. In a preliminary series of force field and combined QM/MM calculations, one of my colleagues managed to compute the detachment of TS793 from the active site cysteine reaching the geometry of the non-covalent complex.[1] The corresponding system comprises the protein and the TS793 molecule with a solvation sphere of about 25Å radius around the inhibitor. Within this structure, the catalytic dyad resides in its zwitterionic form with Cys25 being deprotonated and His162 having protons both at the δ- and ϵ-nitrogen center. The TS793 molecule is located at a distance of 3.2Å measured between the nucleophilic sulfur atom of Cys25 and the C_β atom of the inhibitor. In contrast to the crystal structure, TS793 is not found in the favorable geometry having the intramolecular hydrogen bond. Instead, the aminomethyl substituent is rotated nearly 180°, such that the H29 atom is directed towards the carbonyl oxygen atom of Gly23. The orientation and distance (2.64Å) however indicate, that though stabilizing, this interaction is no real hydrogen bond. A comparison between this complex and the crystal structure of K11777 revealed another interesting feature concerning the orientation of the phenyl ring. Within the enzyme-inhibitor complex of K11777, the phenyl ring is directed away from the protein towards the solvent, probably being stabilized by interactions with the adjacent aromatic ring of the homophenylalanin substituent of K11777. In the complex of TS793 however, the phenyl ring is rotated about 120° around the C_α-S1

[1] unpublished results

bond, such that it lays flat against the protein surface, mainly the aromatic system of Trp184. A similar deviation was already reported between the crystal structure of K11777 and Rhodesain when being compared to crystal structures of cruzain and K11777 or falcipain and K11017, another closely related vinylsulfone with an identical chemical structure in the vicinity of the vinyl scaffold. The orientation of the phenyl ring was observed to correlate with the spatial demands of the side chains of Met145 (Ala166 in cruzain), leaving enough space for the ring to lay flat against the protein within cruzain and falcipain but blocking this position in rhodesain. However, the authors also report, that there might be a significant transition between the two orientations leading to an equal population of the two positions with only one of them being provided in the crystal structure data set.[179] Whereas the steric interactions between the phenyl ring and the Met145 residue can be used to rationalize the preference of either one or the other of these structures, a much stronger impact is expected to arise from the sulfone group itself, forming several hydrogen bonds with the protein. Within the crystal structure 2p7u, the atomic positions indicate three hydrogen bonds being formed between the sulfone oxygen atom O4 and the N-H functions of Gln19, His162 and Trp184. With these three hydrogen bonds, the sulfone group acts as anchor, holding the adjacent vinyl group in a position allowing the covalent attachment to the active site cysteine to happen. Although rotated about 120° with respect to this crystal structure, the formation of these three hydrogen bonds is still possible within the calculated non-covalent complex, thus contributing equally to the energy of either of the orientations of the phenyl ring. However, the distances between the corresponding heavy atoms are significantly larger in the non-covalent complex when being compared to the crystal structure, which indicates a tighter fixation of the covalent adduct and a probably more negative reaction energy.

4.3 Calculations

Based on the introduced above several series of calculations have been conducted. A basic type of calculations is the optimization of the geometry by minimizing the potential energy. Furthermore, as a comparison between the internal influences and those arising from changes in the environment is to be made, the electronic structure has to be evaluated for any of the geometries of TS793 determined for the diverse systems by keeping the geometry fixed but changing the environment to all possible combinations. Once all electron density distributions are obtained, observable quantities can be compared such as the absolute electron density at selected points in space, the molecular dipole moments or the total energy. Also the analysis of the electron density by means of AIM is envisaged. Furthermore, a more qualitative

approach is followed by looking at the electrostatic potential or the d_{norm} parameter and compare them between the individual systems.

4.3.1 Crystal Structure of TS793

For the calculations on the crystal fragment of TS793 a validation of the methods as well as the structure itself was necessary. Once a satisfying quality was reached, the electron densities of various subsystem were computed and analyzed.

Verification of the Force Field

With the crystal fragment constructed from the crystallographic information file using the `Decifer` program, a system for force field calculations was set up. The parameters have been generated via Swissparam[164] with atomic charges taken from AIM calculations on the isolated molecule. As the calculation was primarily intended to verify a reasonably good representation of the system, a minimization of an ellipsoidal region around a central molecule within the crystal fragment was conducted. Whereas any molecules within this region (≈60Å×45Å×45Å) were allowed to relax freely, the surrounding atoms have been rigidly fixed at their initial coordinates derived directly from the crystal structure. With this setup, the flexibility within the active region should be large enough to reveal structural changes due to the force field description while the electrostatic influence of the crystal is assumed to be adequately modeled by the surrounding layers. To avoid issues due to the truncation of coulombic interactions, the calculation was conducted without any cutoff for the electrostatics. Boundary effects originating from the finite crystal fragment used are expected to be negligible due to the large thickness of the layers encompassing the active region and their rigid fixation. The minimization was carried out using the NAMD[29] program because of its beneficial implementation of parallel computations. However, NAMD does not provide a convergence criterion for minimizations but only accepts a predefined number of minimization steps. The number of steps was thus set to 2,000,000 and the convergence was monitored by the convergence of the RMSD value measured along the trajectory. Using the 6728 atoms that were allowed to relax as reference point, the RMSD value in the atomic positions converged to less than 0.7Å within the first 6,000 steps of the minimization, after which the calculation was terminated. The deviations in the positions of the individual atoms were found to differ by up to nearly 1.8Å, however this has been found to only occur within the center of the active region. Furthermore, adjacent atoms move almost in unison therefore resulting in only negligible geometry changes within each single molecule but translations of the molecules as a whole. The largest RMSD values are thus the result from

smaller but accumulated inaccuracies in the description of intermolecular forces, not unexpected for the use of Charmm with a system well outside the structures it was developed for. Internal changes in the geometry have only been observed for the aminomethyl substituent, for which the force field parameters resulted in a slight pyramidalization of the amine. The reason might be the lack of conjugation between the amine and the vinyl scaffold in a coplanar arrangement which can not be properly described within the force field. Despite the issues observed with the position of individual atoms, the geometry of the whole system and especially that of individual molecules remained well within a reasonable range. This also indicates, that the use of AIM charge as atomic charges does not introduce additional artifacts. Therefore, the Charmm force field performed unexpectedly well for the actual system. However, to avoid any biases, subsequent calculations on the crystal fragment have been performed without any relaxation of atoms which are treated by classical mechanics.

QM/MM Optimizations on Monomers and Dimers

The determination of the electronic structure of TS793 within its crystal surrounding requires the use of combined QM/MM methods being able to describe the electron density within the TS793 molecule but still taking the electrostatic interactions with its surrounding into account. However, simple taking the provided crystal structure of TS793 and computing its electronic structure is not sufficient. The electron density distribution within a molecule is very sensitive to the exact geometry of the compound[58] and requires a properly optimized structure at the same level of theory that is used to determine the electron density itself. Hence, optimizations of TS793 and selected dimers had to be conducted using density functional theory (B3-LYP/TZVP) to optimize the corresponding model systems. The environment was kept at its initial geometry such as to introduce as little bias to the system as possible. Its influences onto the model system were captured using the electrostatic embedding. Due to the structure of the crystal fragment, there was no need for covalent bonds to cross the boundaries between QM and MM subsystems thus rendering their special treatment obsolete. All optimization have been verified by numerical frequency calculations, none of which yielded imaginary frequencies.

As all molecules within the crystal structure of TS793 experience exactly the same environment due to their symmetry relations, calculating the optimal geometry for a single monomer was sufficient. As the molecules occur in enantiomeric pairs with respect to the plane of the vinyl scaffold, a monomer was chosen which resembles most closely the geometry of TS793 within the protein complex. The calculation was carried out using the ChemShell package.[51] The changes between the crystal structure and the optimized

monomer geometry are rather low with a displacement of less than 0.5Å for any individual atom and an RMSD value of about 0.2Å averaged over the whole molecule. Whereas bond lengths are hardly affected, the most significant changes are found in the position of the H29 atom, caused by a small rotation of the aminomethyl substituent around the C8-N6 bond. This change however results in an increase of the hydrogen bond between H29 and O3. As the latter is embedded in a hydrogen bond network between two adjacent molecules, another QM/MM optimization was performed in which the QM system was enlarge to encompass both molecules of the dimer. Hence, all four hydrogen bonds between the sulfone and aminomethly substituents of the two molecules are treated on an equal theoretical basis. This is well reflected in the optimized geometry which exhibits an almost complete equilibration of the hydrogen bond lengths between the intra- (2.18Å) and the intermolecular (2.5Å) hydrogen bonds. Interestingly however is that the vinylsulfone scaffolds remain closer to their initial geometry than that of the monomer calculation before. Also the overall RMSD value is reduced to only 0.35Å. The almost identical hydrogen bond lengths seem to stem from a combination of a resonance assisted hydrogen bond within each of the monomers[191,192] and the effect of the dimerization of two such moieties within the crystal. The optimizations of other possible dimers between adjacent TS793 molecules again lead to slightly larger deviations between the optimized geometry and that within the experimental crystal structure (0.5Åand 0.55Å). This holds true especially for the molecule that was relaxed during any of the optimizations described here. The most plausible explanation is the weak interaction between the monomers within the latter two system which results in geometries very close to that of the optimization of a single molecule.

4.3.2 Protein Complexes

For the enzyme-inhibitor complex, the provided structure contains the non-covalently bound TS793 molecule and the active site residues in their zwitterionic, catalytically active protonation state. Yet in order to estimate environmental influences, especially electrostatic interactions, also the neutral form of the two residues is of interest. Although being catalytically active only in the zwitterionic state, the enzyme might reside in its neutral form while being uncomplexed and change into its active form only before or during binding of the inhibitor or even later during the final chemical reaction. Also for the TS793 molecule itself, several possibilities have to be considered. The provided non-covalent complex is formed with the TS793 molecule lacking its intramolecular hydrogen bond. This is a result of the optimization process of the lead structure, for which the aminomethyl function was introduced to maximize the binding affinity via the formation of a hydrogen bond

towards the carbonyl oxygen of Gly23. Whereas this hydrogen bond can readily be established in the covalent complex, it get significantly weakened during the detachment of the inhibitor. Hence, several additional calculations had to be performed to investigate the effect of rotating the aminomethyl substituent and thus the presence or absence of the intramolecular hydrogen bond as well as investigations on the protonation state of the catalytic dyad and possible pathways for the proton transfer between them.

Covalent Complex with high binding affinity

In a preliminary step, the provided complex had to be converted from the Amber force field that was used previously to the Charmm one employed within this work. The parameterization within Charmm was straight forward as the parameters for TS793 had been generated previously for the crystal fragment while the protein and water molecules are contained in the standard parameter set already. Following the translation of coordinates, topologies and parameters, a preliminary optimization of the geometry was performed, again using QM/MM methods. Due to the aim of studying the bond formation with the protein as well as the proton transfer between two amino acid residues, the TS793 as well as the Cys25 and His162 residues together formed the QM subsystem. This first optimization converged within few cycles and the resulting geometry did resemble that of the provided complex almost perfectly.

In a next step, the proton transfer between the His162 and Cys25 residues should be modeled. The corresponding calculations on the reaction path and the corresponding energy profile had been conducted in form of a potential energy surface using the program package described in the programming chapter. The definition of the system was made identically to that of the first optimization. Additional constraints were used as coordinates for controlling the degree of the proton transfer. The selected distances were those between the cysteine sulfur atom and the transferred proton as well as the distance between this proton and the N_δ atom of His162. Using a second coordinate for the transfer allowed a mode precise localization of the transition state and thus a more reliable activation barrier for the proton transfer reaction. The calculation of the reaction path succeeded without any technical difficulties while the energy profile revealed two clearly defined minima, one for each of the proteins protonation states. While the geometries of the two minima are as expected very similar to one another, the negligible energy difference between them was not. With the geometrical and energetic similarity between the two minima and the only moving particle being a proton, also the activation barrier is very low ($\approx 8 - 10 \frac{kJ}{mol}$).

Additionally to the calculations on the proton transfer reaction, also the bond formation process with the accompanying protonation of the inhibitor

was modeled using the same model system. The constraints used were changed to the distance between the nucleophilic sulfur atoms of the Cys25 residue and the C_β atom of TS793 and the second distance between the C_α atom and the proton on His162. Whereas the activation barrier for this reaction was of lesser interest, the energy difference between non-covalent and covalent complex was. It was determined to about $+60\frac{kJ}{mol}$, being clearly endothermic and in good agreement with the estimations from the preceding QM calculations. The geometry of the covalent complex was found to be very similar to that of the K11777 inhibitor within the crystal structure, except for the rotated phenyl substituent mentioned earlier. A comparison of the hydrogen bond network between the sulfone and the adjacent amino acid residues revealed a much better resemblance of the bond distances compared to the non-covalent complex. The corresponding heavy atom distances are given in Table4.2.

system		O4-N_{His162}	O4-N_{Trp184}	O4-N_{Gln19}
2p7u[179]		3.54	3.15	2.89
covalent	no H-bond	3.88	3.47	2.82
covalent	H-bond	3.62	3.12	2.85
non-covalent	no H-bond; zwitterion	4.14	4.21	3.43
non-covalent	no H-bond; neutral	4.40	4.61	3.36
non-covalent	H-bond; zwitterion	4.61	4.79	3.56
non-covalent	H-bond; neutral	4.54	4.83	3.52

Table 4.2: Heavy atom distances measured between the heavy atoms of the sulfone and the adjacent amino acid residues participating in the hydrogen bond network anchoring the sulfone moiety to the active site of Rhodesain. All distances are given in Å.

Covalent Complex with Intramolecular Hydrogen Bond

Whereas the TS793 appears to be suitably stabilized within the covalent complex due to the hydrogen bond formed between its H29 atom and the Gly23 residue, a direct comparison with the crystal structure is not possible as the geometries differ too largely from each other. Hence, the aminomethyl substituent was rotated within the non-covalent complex by $\approx 180°$ such that the formation of the intramolecular hydrogen bond is possible again. This was performed on both non-covalent complexes such that the two protonation states of the enzyme are taken into account for. An unconstrained minimization of both structures was conducted followed by another reaction path calculation for the proton transfer reaction. As expected, the

formation of the intramolecular hydrogen bond has a large stabilizing effect, lowering the energies of the non-covalent complexes where it is present by about $-35\frac{kJ}{mol}$ with respect to those that lack this bond. While the difference in energy between the two protonation states is still very low, in the presence of the intramolecular hydrogen bond the zwitterionic state is favored by $-4\frac{kJ}{mol}$. Depending on the direction of the proton transfer, the activation barrier is $\approx 6-10\frac{kJ}{mol}$. Compared to the non-covalent complexes lacking the hydrogen bond, the interactions between enzyme and inhibitor appear to be weakened as the heavy atom distances between the sulfone oxygen atom and the adjacent heavy atoms of the hydrogen bond donor residues are increased by about 0.2Å. With the hydrogen bond formed between the sulfone and the aminomethyl substituent of TS793, a second reaction path for a potential bond formation between Rhodesain and TS793 was computed. Except for the modified starting geometries, the reaction path was approximated identically to the one computed before. The obtained geometry for the covalent enzyme-inhibitor complex resembles that of the other covalent structure quite well with the only exception being the rotated aminomethly group. Especially in the position of the amino acid residues that constitute the active site region, only marginal deviation can be observed. Interestingly, despite the large differences in the aminomethyl group and corresponding hydrogen bonds, the relative energies of the two covalent enzyme-inhibitor complexes are almost identical. Together with the non-covalent complexes, this results in three energy levels. The most favorable structures are non-covalent complexes in which the intramolecular hydrogen bond of TS793 is present. The protonation state of the enzyme hardly affects the energy being only slightly in favor for the zwitterionic one. The next energy level ($\approx +35\frac{kJ}{mol}$) comprises the non-covalent complexes in which TS793 lacks its hydrogen bond and thus its stabilization. The highest energy level ($\approx +95\frac{kJ}{mol}$) contains the two covalent complexes as products of the addition reaction.

4.3.3 Electron Density Determination

In order to separate the internal and environmental influences on the electron density of TS793, several new subsystems had to be set up. In order to remove effects that arise from the molecular geometry, calculations have been performed on TS793 in various environments but using the same geometry. For example taking the geometry of TS793 from the optimized crystal fragment, the electron density distribution was determined under consideration of the surrounding via a field of point charges, neglecting any environmental effects by performing a vacuum calculation as well as approximating liquid aqueous surrounding by using the COSMO. Whereas these systems could readily be derived from the QM/MM calculations on the crystal fragment, the non-covalent protein complexes needed some further preparation. During

the corresponding QM/MM calculations, not only the TS793 molecule but also the Cys25 and His162 residues were treated by using quantum mechanics. Nevertheless, being only interested in the electron density of TS793, the QM system had to be reduced to only that molecule. The Cys25 and His162 residues were thus incorporated into the field of point charges around TS793 by placing the atomic charges from the force field parameters at the corresponding coordinates. For an easier reference to the large set of subsystems, labels were assigned according to Table 4.1.

	environment for electron density determination			
geometry optimization	crystal	protein	vacuum	solution
crystal	cc	-	cv	cs
protein	-	pp	pv	ps
vacuum	-	-	vv	vs
solution	-	-	sv	ss

Table 4.3: Table of the subsystems derived from QM/MM calculations on the crystal fragment or the protein complexes. The first letter denotes the model system used to optimize the geometry whereas the second one indicates the environment used for the electron density determination only.

In addition to calculations using only the TS793 molecule to determine its electron density, further computations have been performed for the dimer system in the crystal fragment as well as larger systems comprising the TS793 molecule and all adjacent molecules or residues. The latter are then used for the AIM analysis of the interaction pattern, especially in its differences between the crystal fragment and the protein complexes as well as among these complexes.

4.3.4 Molecular Dipole Moments

A quantity, that can readily be extracted from quantum mechanical calculations is the molecular dipole moment μ. It can be used as a measure for the polarization of the electron density. Furthermore, the dipole moments provide the possibility to verify the correlation between the actual electron density determinations and the latter AIM analysis, as dipole moments could be derived from both types of calculations. Fortunately, the dipole moments derived from either sort of calculations were virtually identical, thus only a single value for each system is given below.

Most obviously from Figure 4.9 is that the computed dipole moments are indeed highly sensitive to the surrounding. Values computed for the isolated

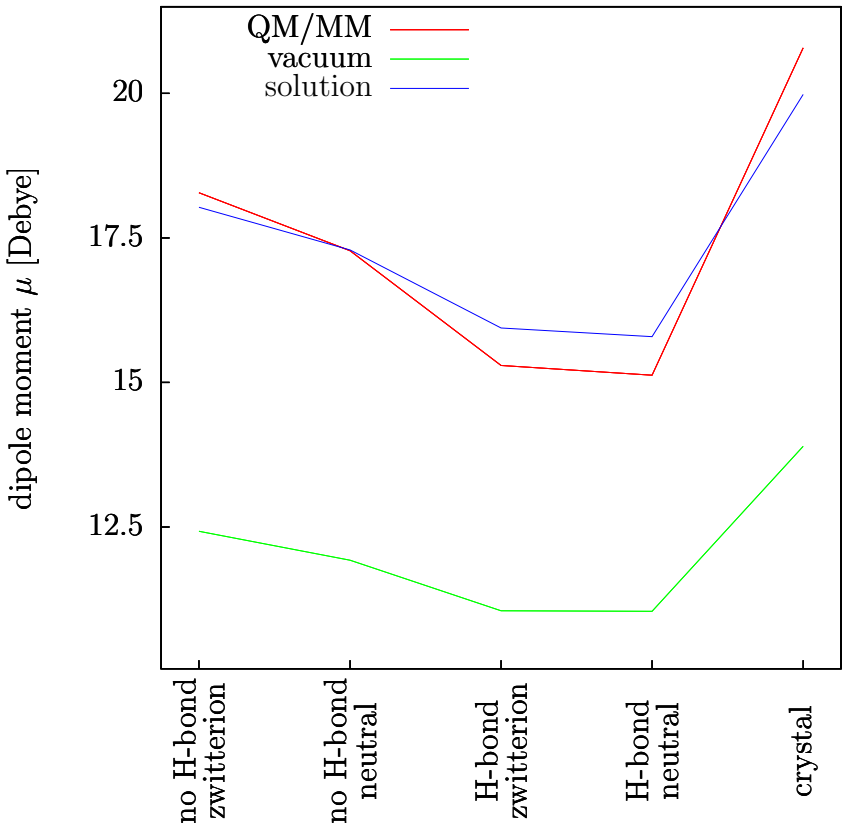


Figure 4.9: Molecular dipole moments calculated for TS793 within the four non-covalent enzyme-inhibitor complexes as well as within the crystal fragment. Each curve represents one of the three possible environments, explicit modeling of all surrounding atoms by QM/MM, vacuum calculations without any environmental influences and aqueous solution modeled with COMSO. All dipole moments are given in Debye.

molecule are about 30% lower than those obtained within the QM/MM calculations or those employing COSMO. This trend is consistent throughout the complete series. Concentrating on the dipole moments computed in vacuum, one can clearly see the influence of the molecular geometry on μ. Among the protein-derived systems, the intramolecular hydrogen bond seem to decrease the polarization of the electron density with the dipole moment being about 1 to 1.5 Debye below those of geometries without the H-bond. The geometry of the crystal structure however also possesses the intramolecular hydrogen bond, yet its dipole moment exceeds that of any other system by at least 2 Debye. The reason for that was found in separating the dipole moment of the whole molecule into its atomic contributions. Based on the AIM calculations, the largest contributions to μ_{total} originate from the oxygen atoms of the sulfone and the nitrogen atom of the cyanide substituent. Further contributions results from the atomic dipole moments of the vinyl carbon atoms, the second nitrogen center and the thiomethyl sulfur atom. Although being almost equal in their absolute values between the protein and crystal derived systems, the orientations of these contributions differ significantly. Especially the orientation of the thiomethyl substituent is nearly reversed, thus adding up with that of the sulfone within the crystal geometry but decreasing that of the protein derived systems.

Going from the vacuum systems to those using COMSO, a large increase in the molecular dipole moments can be observed. This is in good agreement with the expectations. The COSMO works by placing small charges around the computed system which stabilize a charge separation within the molecule. Hence, using COSMO in the calculations, all structures are expected to exhibit a larger polarization and, concerning the energies, the more polar a structure is, the larger the energetic stabilization can be. For the investigated system, this trend could clearly be found. Whereas the trend in the computed dipole moments for the five systems did not change, the spread of values increase spanning a range of more than 4 Debye, whereas that of the vacuum system had been determined to be well below 3 Debye.

The third curve depicted in Figure 4.9 represents the dipole moments of TS793 calculated with an explicit modeling of the crystal or protein surrounding. Although being conceptually different to the two previous series, the trend of dipole moments still remains the same as before. A comparison between the curve of the values calculated using COSMO with that considering the explicit environment exhibits an astonishingly good correspondence. Whereas the COSMO is effectively developed to stabilize polar structures, both the protein and crystal surrounding seem to have a comparable effect. Especially for the crystal structure, the dipole moment even exceeds that obtained with COSMO, indicating that the stabilization of large dipole moments within this system is considerably large. Similar finding are reported for several other crystal structures as well,[193,194] although the stabilization

does not seem to be a universal property of crystals in general.

4.4 AIM Analysis

The AIM calculations were started from wavefunction files (*.wfn) that contain the analytical solution to the electron densities optimized using the Turbomole package. For the conversion, the tm2molden and the molden2aim tools have been used. The program for the AIM computations that has been used in this work is AIMAll.[195] To validate some of the results, further program have also been used such as AIM2000[196] or MultiWFN.[197] However, all results shown below result from calculations with AIMAll.

As the strongest interactions of TS793 with its surrounding were expected to be found in the electron densities of the sulfone moiety, the corresponding S=O double bonds have been studied. However, the differences in the electron densities within each of these bonds as well as between them are below $0.01\frac{e}{a^3}$. Even the laplacian values hardly differ with a maximum deviation of less than $0.2\frac{e}{a^5}$. Similar findings have been made for any other bond within TS793, except for the intramolecular hydrogen bond betwewen S1 and H29.

4.4.1 Intramolecular hydrogen bond

To give an estimate on the strength of the intramolecular hydrogen bond being formed between the H29 atom and one of the two sulfone oxygen atoms, the corresponding systems except the two protein complexes lacking this bond have been analyzed. For the comparison, the electron density ρ has been measured at the bond critical points of the hydrogen bond $\rho_{BCP}(OH)$ as well as the BCP of the bond between the amine nitrogen N6 and the proton $\rho_{BCP}(NH)$. The measured values are given in Table 4.4.

Most obviously, the density measured along the N-H bond, constituting a usual, covalent bond, is roughly ten times as large as that determined for the O$\cdots$H bond. This trend could be observed throughout all investigated systems. Concentrating on the values of the hydrogen bond, the solution model and the protein complexes agree well with each other, while the crystal fragment and the vacuum calculations yield significant different values. As one could expect from the vacuum system, the only one having no environment around the TS793 molecule at all, the hydrogen bond is found to be very strong with about 20% more charge accumulation at the BCP compared to the protein and solution models and almost three times as much as that found in the crystal fragment. The large value originates from the hydrogen bond being the only stabilizing effect within the vacuum model which thus is a dominating factor. Within all other systems, the strong positive polarization at the N-H function and the negative polarization of the

O4/O4 $\cdots$ H29 bond			
system	$\rho_{BCP}(OH)$	$\nabla^2\rho(OH)$	$\epsilon(OH)$
global optimum in vacuum	0.0350	0.1237	0.0669
global optimum with COSMO	0.0278	0.1060	0.0591
crystal fragment	0.0123	0.0524	0.6705
protein, zwitterion	0.0278	0.1055	0.0306
protein, neutral	0.0291	0.1087	0.0563
N6 - H29 bond			
system	$\rho_{BCP}(NH)$	$\nabla^2\rho(NH)$	$\epsilon(NH)$
global optimum in vacuum	0.3328	-1.8383	0.0438
global optimum with COSMO	0.3358	-1.8679	0.0404
crystal fragment	0.3174	-1.7473	0.0385
protein, zwitterion	0.3295	-1.8398	0.0382
protein, neutral	0.3338	-1.8693	0.0396

Table 4.4: Values determined via AIM analysis at the bond critical points of the two bonds with H29 and adjacent heavy atoms. Densities ($\frac{e}{a^3}$) and ∇^2 values ($\frac{e}{a^5}$) are given in atomic units, bond ellipticities as dimensionless numbers.

sulfone group get significantly reduced due to a stabilization by neighboring residues or the solvent model. In the crystal model however, the measured density at the BCP(OH) is unexpectedly small, indicating a much stronger interaction of the N-H and sulfone groups with the environment than within any other system. The reason can be found in the formation of dimeric structures within the crystal fragment in the shape sketched in 4.8. For each H29 atom of either monomer, not only one but two chemically identical oxygen atoms are available, sharing the stabilizing contributions and thus the electron density. This effect even extends to the density between the H29 and the N6 atom, for which nearly identical densities are found in all systems, except for the crystal model. Here, the density is reduced by up to 5% compared to the four other model systems. The decrease in $\rho_{BCP}(NH)$ for the crystal system is also reflected in the ∇^2 value being the least negative among all systems indicating a smaller charge accumulation at the BCP. For $\rho_{BCP}(OH)$ however, the ∇^2 values in contrast to the lowered electron density. A closer look at the density distribution around the BCP revealed, that although $\rho_{BCP}(OH)$ is reduced, an increase in the density above and below the plane of the two H29 and the sulfone oxygen atoms could be found. This π-like distribution is also reflected in the bond ellipticites ϵ, which show almost perfectly rotational symmetry for the hydrogen bond in vacuum, solution or the protein complex, but a significant distortion within

the crystal structure. The values of $\epsilon_{BCP}(OH) = 0.6705$ even exceeds that of prototypical carbon=carbon double bonds which usually reach up to only ≈ 0.45. The reason for the extraordinary electron distribution around the hydrogen bond within the crystal fragment arises from the unique chemical substructure which not only comprises a quadrilateral shape of two donors and two acceptors, but which is in conjugation with the vinyl scaffold of both monomers and, at least partially, the sulfone groups. Within each monomer, the hydrogen bond can be interpreted as an example of the resonance assisted hydrogen bonds reported by Gilli et al.[191,192] or Grabowski et al.[198] Hydrogen bonds of this type are generally extraordinarily strong and exhibit a significant degree of π-character as measured by the ϵ value for the current system.

Interactions in the crystal

Compared to the hydrogen bond network between dimers of TS793 within the crystal fragment, the interactions between adjacent molecules are much weaker. Although further bond paths could be found, especially between the hetero atoms of neighboring molecules, the low densities at the corresponding BCPs hardly exceed a value of $0.005\frac{e}{a^3}$, the value usually used to determine the outer sphere of a molecule.[24] Probably due to their proximity, a bond path could be found between the oxygen atom, that does not participate in the hydrogen bond network and one of the hydrogen atoms of the methylamine substituent. However, this interaction could hardly be characterized as hydrogen bond and originates mainly from the geometry and the opposite polarization of the sulfone and the terminal methyl group. Similarly, BCPs are found between the S2 atom and the N5 and O3 atoms participating in the hydrogen bond network. Although being slightly above the value of $0.005\frac{e}{a^3}$, these interactions are also mostly electrostatic in nature.

Interactions in the protein

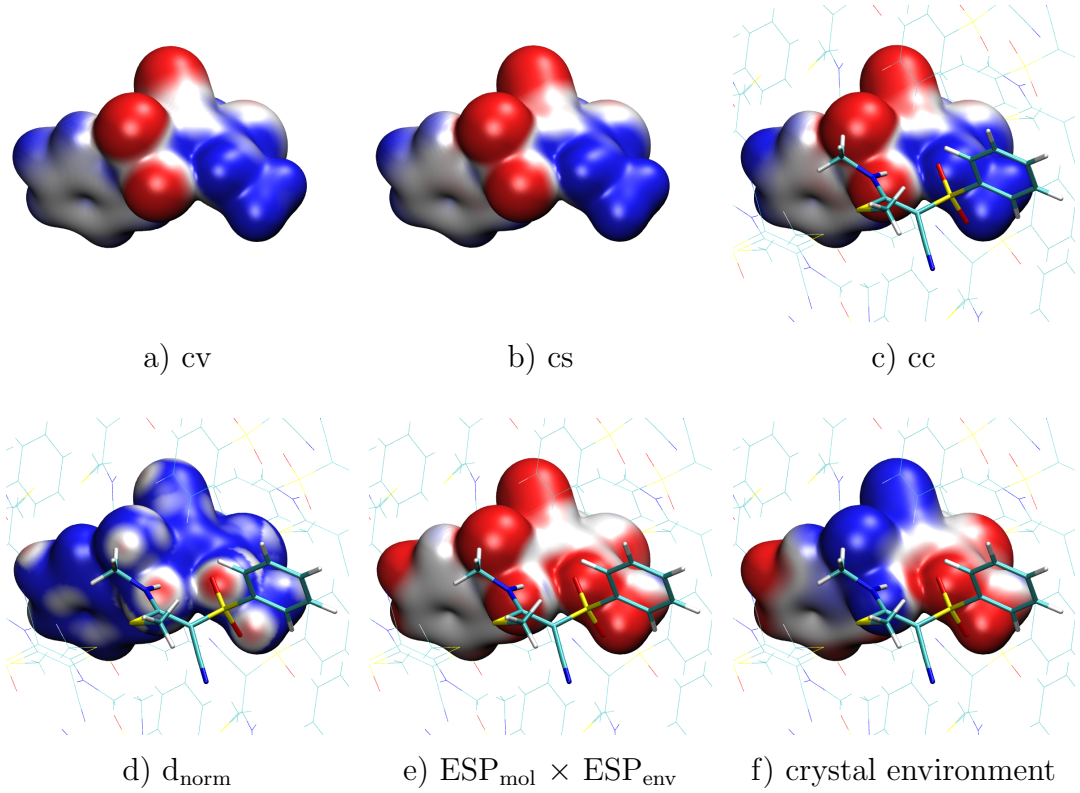
Within the protein models, the most significant difference is of cause the presence or absence of the intramolecular hydrogen bond in TS793. However, further bond paths with significant electron density at their BCPs could be found. Most of these paths are located around the oxygen atoms of the sulfone, which form several hydrogen bonds with adjacent water molecules. Similarly, the cyanide substituent accepts a hydrogen bond from one of the water molecules. According to ρ_{BCP} for these hydrogen bonds, they are throughout stronger than the interactions between the O4 atom and the Gln19, His162 and Trp184 residues which are supposed to anchor the inhibitor within the active site. In fact, no path directly connecting O4 and any of the proton donor groups could be found in all four systems. Likewise,

non of the non-covalent complexes exhibits a bond path between the Cys25 residue and the C_β atom of TS793. The lack of such a path indicates the necessity of significant rearrangements during the bond formation reaction and partially explains the endothermicity of the bonding reaction.

4.5 Molecular Surfaces

In addition to the investigations carried out using the AIM theory, the computed electron densities were compared in a visual inspection of their three dimensional shape. Therefore, the molecular surfaces were drawn as isosurfaces of $0.005\frac{e}{a^3}$, a value with closely resembles that of the van der Waals radii[63] or Hirshfeld surfaces which are used in crystallography.[71] The benefit of using isodensity surfaces is that they, contrary to van der Waals radii, are able to reflect deformations originating from the chemical bonding within the molecule and, in difference to Hirshfeld surfaces, can also be computed for isolated molecules lacking and explicit surrounding. In Figure 4.10, several such surfaces are shown for TS793 within the crystal fragment and model systems derived from it.

The top row of this graphic shows the electron densities computed for TS793 at the geometry within the crystal fragment but using different environments. The cv system is used as a reference, as it depicts the electron density solely arising from the molecular geometry. Also computed within cv, the electrostatic potential is mapped onto the surface. Displayed in this manner, the graphic is expected to represent as good as possible the major criteria that play a role in molecular recognition or intermolecular interactions in general. For the TS793, the graphic clearly exhibits three positions of negative polarization at the O3, O4 and N5 atoms, a more widespread positive polarization of the thiomethyl substituent and the remainder of the surface without specifically large charge accumulation. A comparison with the cs model shows generally similar features whereas the absolute magnitude of polarization within the cs model clearly exhibits that of the cv model. A similarly large polarization can also be found in the cc system, which very closely resembles the cs system. The difference between the explicit modeling of the crystal environment in cc and the effect of COSMO in cs appear to yield almost identical electron density distributions. One should, however not forget, that an optimization of the molecular geometry with COSMO would most probably also result in a change in the electron density. However, as the geometries are not too different, the solution model represents already a rather good approximation to the full QM/MM calculation. Besides the electrostatic potential of the TS793 molecule itself, the QM/MM calculations also provided to possibility to calculate that of the environment. It is given in Figure 4.10f. Comparing the electrostatic potential of TS793



a) cv b) cs c) cc

d) d_{norm} e) $ESP_{mol} \times ESP_{env}$ f) crystal environment

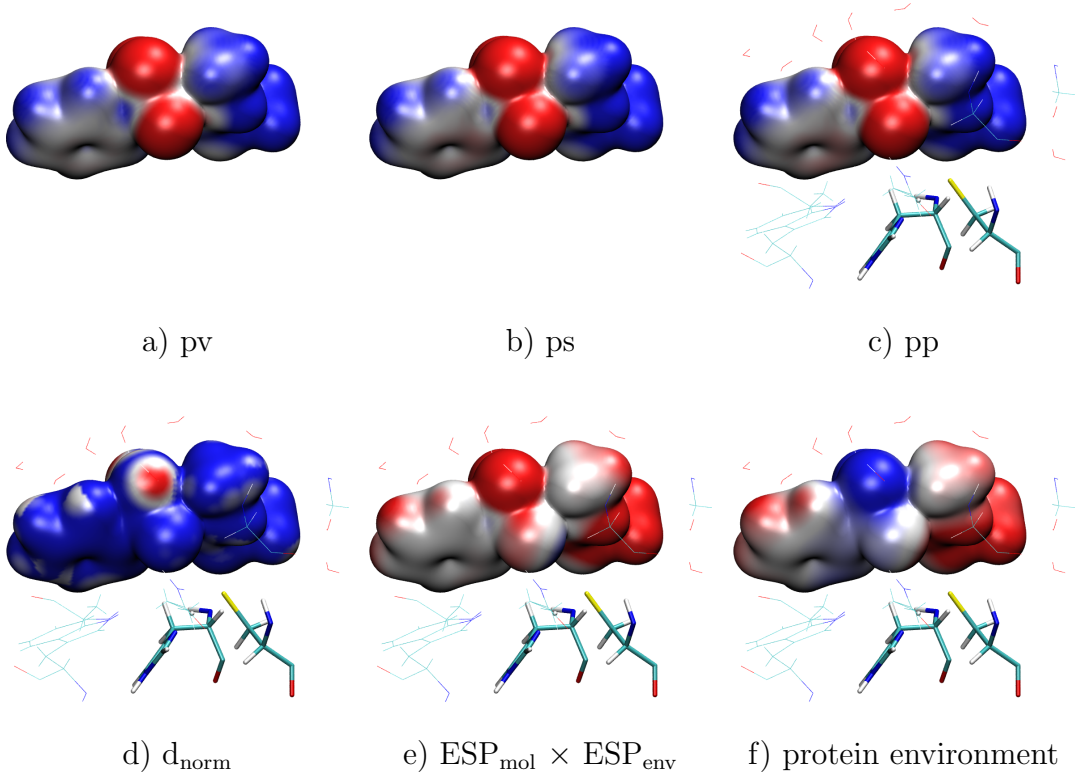
Figure 4.10: The top row (a to c) shows the isodensity surface of the electron density (0.05a.u.) computed at the optimized geometry within the crystal fragment. The color coding, red for negative and blue for positive values, represents the electrostatic potential (±0.025a.u.) arising solely from the molecule itself. Subgraphic f depicts the corresponding potential originating from the crystal environment. The geometric parameter d_{norm} (±0.5) is depicted in d and e visualizes the product of the electrostatic potential of TS793 with that of its environment (±0.0025a.u.).

with that of the crystal surrounding, one finds an almost perfect complementary between the two. This is even more obvious in Figure 4.10e, which represents the product of the internal and external potentials. Regions of opposite sign result in attraction between the components and are displayed in red. Unpolarized regions result in small values, colored white whereas an equal sign in the electrostatic potential would be display in blue. However, such areas could not be found within the crystal structure, indicating that potential repulsive forces are not present. Hence, both molecules with each dimer can approach each other very closely, which can be seen in Figure 4.10d. The d_{norm} parameter being mapped onto the surface provides information on the proximity of two molecules with respect to their vdW radii. In a simplified fashion, white regions in d_{norm} represent exact vdW distances. Blue regions indicate larger separations and red regions are found, where particles approach each other more than their vdW radii would allow them to do. The latter can be found at the dimer interface and correspond well with the hydrogen bond network and the, obviously sterically determined, interaction of the thiomethly substituent with the second oxygen atom of the sulfone.

For the protein complexes, similar graphics have been prepared which are given in Figures 4.11, 4.12, 4.13 and 4.14.

Figure 4.11 was derived from the protein complex having no H-bond within the inhibitor and the catalytic dayd in its zwitterionic state. As indicated in the top row, the same trend as that within the crystal fragment could be observed with least polarization in the pv system and significant larger polarizations in the other two. Again, the solution model can reproduce the electron density distribution and the electrostatic potential of the full QM/MM calculation rather well. As before, the electrostatic potential of the molecule as compared to that of the environment and again a high complementary was expected. Whereas this could be found at the upper half of the molecule and the thiomethyl substituent on the right side, the correspondence between the potentials between the TS793 molecule and the catalytic dyad is very low. Furthermore, as depicted in Figure 4.11e, the shape of the potentials is such, that considerable repulsion could be found between the sulfone and the Cys25 sulfur atom.

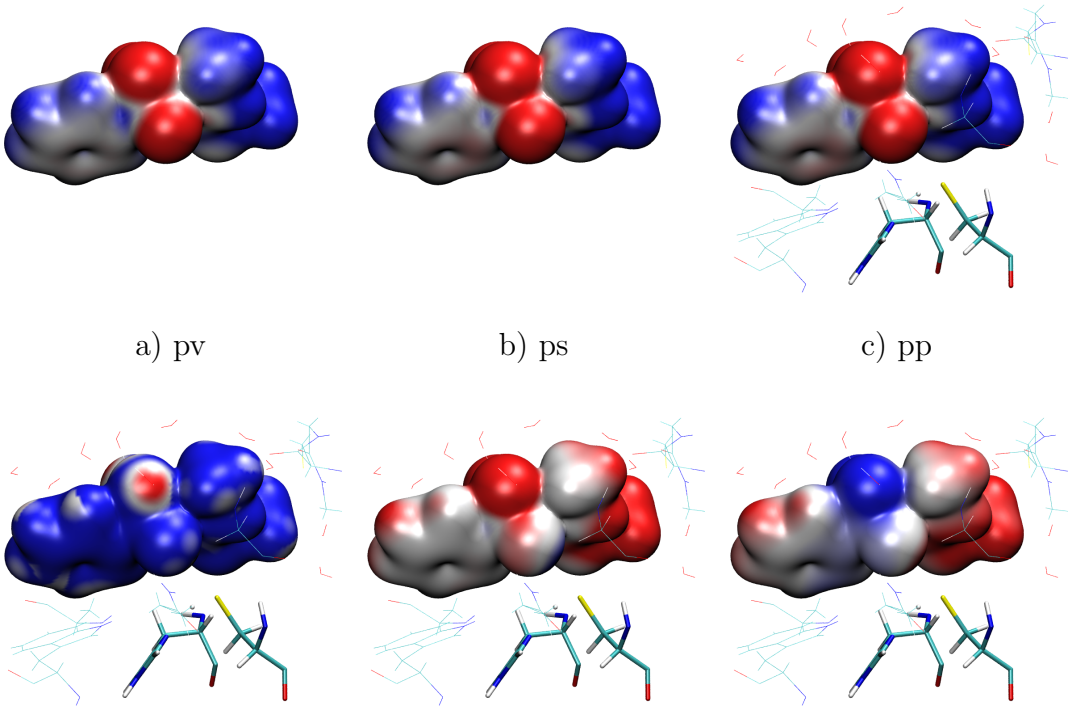
The repulsion is caused by the negative polarization on both, the inhibitors oxygen atom and the sulfur atom of Cys25. Thus the same type of graphics was also prepared for the neutral protonation state of the enzyme, which could potentially lower the repelling force. Whereas this can in principal be found in comparing Figure 4.11 and 4.12, the differences are only marginally. As the hydrogen atom is only shifted by about 0.6Å between the two system, changes in the electrostatic potential are also very small. Furthermore, the hydrogen atoms is moved almost linearly between the His162 and the Cys25 residue such that the large size of the sulfur atom can already screen much of



a) pv b) ps c) pp

d) d_{norm} e) $ESP_{mol} \times ESP_{env}$ f) protein environment

Figure 4.11: The top row (a to c) shows the isodensity surface of the electron density (0.05a.u.) computed at the optimized geometry within the non-covalent complex (no H-bond, zwitterion). The color coding, red for negative and blue for positive values, represents the electrostatic potential (±0.025a.u.) arising solely from the molecule itself. Subgraphic f depicts the corresponding potential originating from the protein environment. The geometric parameter d_{norm} (±0.5) is depicted in d and e visualizes the product of the electrostatic potential of TS793 with that of its environment (±0.0025a.u.).

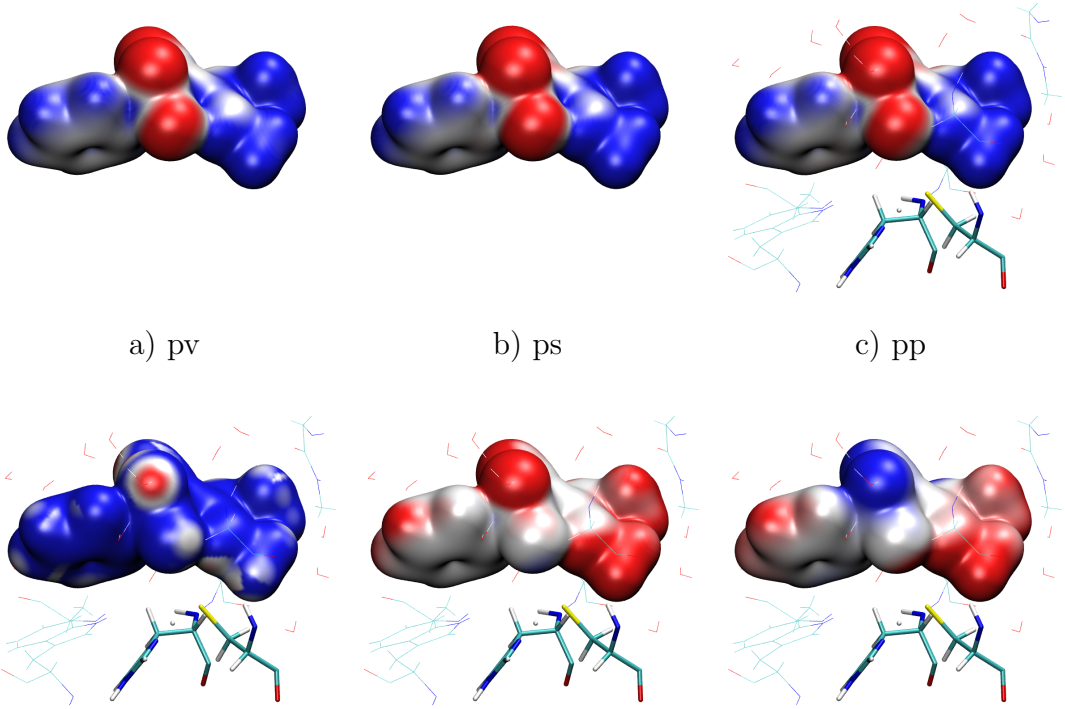


a) pv b) ps c) pp

d) d_{norm} e) $ESP_{mol} \times ESP_{env}$ f) protein environment

Figure 4.12: The top row (a to c) shows the isodensity surface of the electron density (0.05a.u.) computed at the optimized geometry within the non-covalent complex (no H-bond, neutral). The color coding, red for negative and blue for positive values, represents the electrostatic potential (±0.025a.u.) arising solely from the molecule itself. Subgraphic f depicts the corresponding potential originating from the protein environment. The geometric parameter d_{norm} (±0.5) is depicted in d and e visualizes the product of the electrostatic potential of TS793 with that of its environment (±0.0025a.u.).

the hydrogen's positive charge. The repelling force between the Cys25 sulfur atom and the inhibitor can also be found in the d_{norm} parameter which in both graphics exhibits almost solely regions of distances above the vdW contact distances (blue color). The only regions of closer proximity between TS793 and its surrounding are found on the upper half of the molecule, where several hydrogen bonds are formed with solvent molecules.



a) pv b) ps c) pp

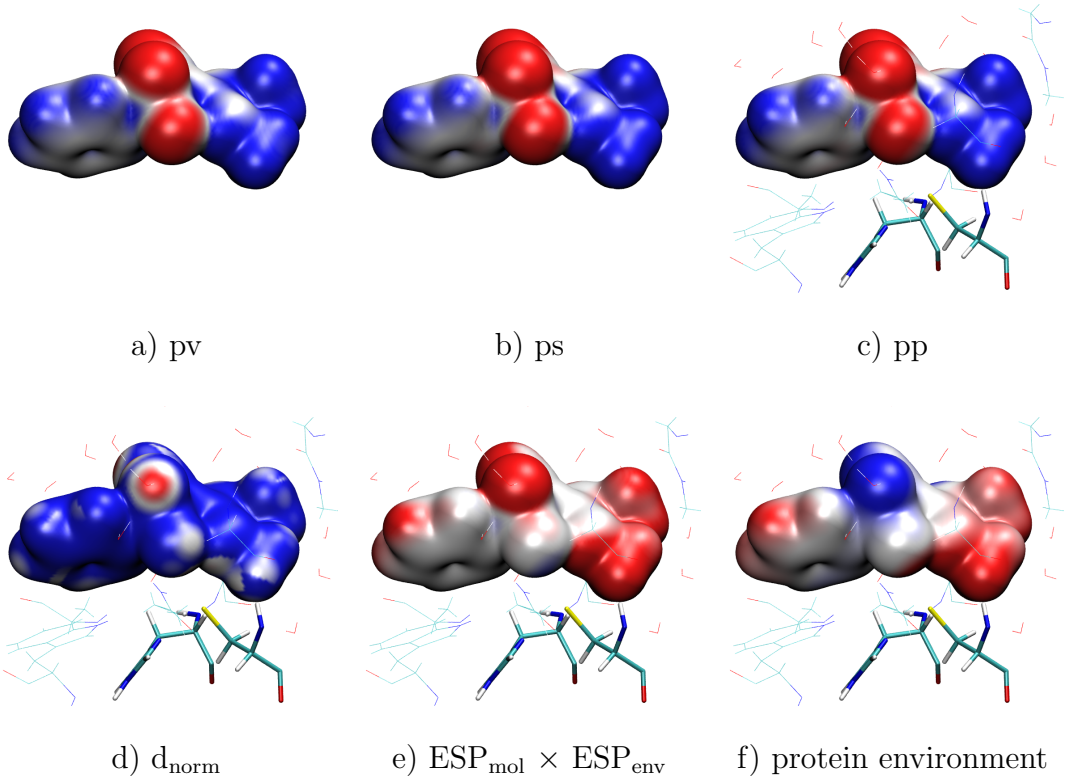
d) d_{norm} e) $ESP_{mol} \times ESP_{env}$ f) protein environment

Figure 4.13: The top row (a to c) shows the isodensity surface of the electron density (0.05a.u.) computed at the optimized geometry within the non-covalent complex (H-bond, zwitterion). The color coding, red for negative and blue for positive values, represents the electrostatic potential (±0.025a.u.) arising solely from the molecule itself. Subgraphic f depicts the corresponding potential originating from the protein environment. The geometric parameter d_{norm} (±0.5) is depicted in d and e visualizes the product of the electrostatic potential of TS793 with that of its environment (±0.0025a.u.).

Going from the protein complexes in which the TS793 lacks the intramolecular hydrogen bond (Figures 4.11 and 4.12) to thus that do have it (Figures

4.13 and 4.14), the geometry of the molecule becomes significantly more similar to that found in the crystal fragment. This can indeed be confirmed by comparing the shape of the densities drawn in Figure 4.10 with those of Figures 4.13 and 4.14. Whereas steric interactions between the terminal methyl function of the aminomethyl substituent causes the former to arrange slightly above the plane of the vinyl scaffold, the formation of the hydrogen bond between H29 and O4 removes this tension and enables a favorable orientation of the methyl group slightly below this plane. By this rotation however, the methyl group approaches the sulfur atom of Cys25 creating sterical repulsion between the TS793 molecule and the cysteine. Upon minimization of the energy, the systems responds with the TS793 molecule moving a bit further away from the catalytic dyad. The distance between the nucleophilic sulfur atom of Cys25 and the C_β atom of the inhibitor for example increases from about 4.0Å within the complexes lacking the hydrogen bond to about 4.3Å when the hydrogen bond is formed. This increase distance partially explains the larger reaction energies when the reaction is started from the TS793 molecule having the intramolecular hydrogen bond. Besides the smaller changes in the geometrical arrangement, the complementary between the molecules electrostatic potential and that of the protein surrounding are comparable among all four protein complexes. Especially on the outer perimeter of the molecule, hardly any changes could be found. The rotation of the aminomethyl function however is also reflected in the electrostatic interaction. Whereas without the intramolecular hydrogen bond only the relatively weakly polarized methyl function comes into contact with the protein, the H29 atom is too far away from any of the proteins residues for a significant interaction. Upon formation of the H-bond however, the slightly positively polarized methyl group comes closer to the sulfur atom of Cys25, resulting in a considerable attraction. The interaction of these oppositely polarized moieties can clearly be identified in Figures 4.13e and 4.14e. The backbone amide N-H function shown in the foreground of these graphics however is again too far away from the methyl group for a notable interaction.

The most significant feature in Figures 4.13e and 4.14e is however the notable repulsion between the sulfone oxygen and the Cys25 sulfur atom. Even in the neutral state of the catalytic residues, the charge of the proton does not compensate for the negative charge on the Cys25 sulfur. Furthermore, due to the change in the position of TS793, this sulfur atom is closer to the sulfone compared to the protein complexes without the hydrogen bond. Hence, the nucleophilic attack on C_β is further hindered.



a) pv b) ps c) pp

d) d_{norm} e) $ESP_{mol} \times ESP_{env}$ f) protein environment

Figure 4.14: The top row (a to c) shows the isodensity surface of the electron density (0.05a.u.) computed at the optimized geometry within the non-covalent complex (H-bond, neutral). The color coding, red for negative and blue for positive values, represents the electrostatic potential (±0.025a.u.) arising solely from the molecule itself. Subgraphic f depicts the corresponding potential originating from the protein environment. The geometric parameter d_{norm} (±0.5) is depicted in d and e visualizes the product of the electrostatic potential of TS793 with that of its environment (±0.0025a.u.).

4.6 Conclusions

The TS793 molecule was taken as a model system to investigate the environmental influences on its electron densitiy arising from the target protein it was designed for, the surrounding of the crystal structure of the pure compound as well as that of aqueous solution. Additional reference calculations neglecting any environmental influences have also been carried out. The structure of the model compound was optimized with respect to the various surroundings resulting in a model for the crystal fragment, four non-covalent protein complexes and the two reference systems for solution and vacuum. During the optimizations, estimations on the reactivity were made which are in good agreement with pure quantum mechanical calculations carried out earlier.[180] For the crystal structure as well as for the protein complexes, structures were found that resemble well the energetically most favorable geometry of TS793. The interactions with the environment however were found to significantly differ from each other, as the formation of dimers within the crystal has a largely stabilizing effect on the whole system whereas the newly introduced aminomethyl substituent appears to destabilize potential complexes with the enzyme. Experiments on the inhibition potency of TS793 against Rhodesain even indicate, that the changes in the molecular structure from the K11777 inhibitor to TS793 cause the molecule to no longer address the active site, making a covalent inhibition impossible.[2]

[2] unpublished results

5 Programming

5.1 Two-Dimensional RMSD Measurements

Simulations on the dynamics of molecular systems are usually carried out to obtain data on the statistical behavior of the system. Of particular interest are the mean value of a measured quantity as well as its fluctuation and evolution with time. Basic visualization tool are implemented in many molecular visualization tools such as VMD[28], which allow to plot the value of internal coordinates along the trajectory. A more global measure is the root mean square deviation (RMSD) in the atomic positions which provides information on the deformation of a geometry or the stability of a system. For molecular dynamics of protein systems, a common choice for the reference geometry may be the crystal structure so that the RMSD provides information on the structural deviation hereof. Nevertheless, simply computing the deviations according to the formula above will most probably not result in the minimal RMSD values, as translations and rotations of the system as a whole are still included. Being only interested in the internal deviations, the two geometries have to be aligned to each other before computing the RMSD. There are two algorithms being used to align geometries in visualization programs, the Kabsch algorithm[199,200] and the rotation via quaternions.[201] The former is also implemented in VMD allowing to calculate minimal RMSD values. Nevertheless, analyzing a trajectory by means of these RMSD values suffers from the choice of a single reference structure. Whereas low values indicate a good agreement with the reference, large values of two other geometries do neither imply a structural difference between the two, nor do similar values measured for these geometries indicate any similarity between them. To address this issue, it it necessary to calculate the RMSD values not with respect to a single reference structure, but in a pairwise manner leading to the common two dimensional RMSD plots. Whereas these plots provide detailed information on the mutually pairwise deviations between two geometries, they usually do not represent minimal RMSD values. In most cases, the alignment of all geometries along the trajectory is still performed using only a single reference. While this allows to decrease the computational demands significantly, geometries different from the reference are not necessarily optimally aligned to each other. Furthermore, 2D plots of RMSD values are inherently symmetric with respect to the diagonal, as the comparison of geometry m and n will

yield the same values as that of n with m. While this allows again a reduction in the computations, the corresponding graphs still contain redundant information. For these reasons, the DCDRMSD2D program was developed. This program performs a mutual alignment of each pair of geometries before computing the corresponding RMSD value. To further extend its capabilities, the alignment is carried out either on the entire system or just on a user specified selection of atoms. Thus it is possible to align two geometries for example by the backbone atoms of the protein chain which, due to their higher stiffness, usually gives better results than including the more flexible side chains. In contrast to existing programs, the DCDRMSD2D program accepts a second selection of atoms, for which an additional RMSD measurement is performed. While the first one, alignment and measurement, is stored in the lower half of the 2D plot, the second one, only measurement, is stored above (see Figure 5.1). Thus, the DCDRMSD2D program does not only provide true minimal RMSD values for each pair of geometries but also removes the redundancy of common 2D RMSD plots by providing additional information.

The DCDRMSD2D program accepts the set of geometries in form of a dcd file, e.g. produced by NAMD during a molecular dynamics simulation. The user then has to specify the index number of the atoms which should be used to align the geometries as well as the second set of indices needed for the additional measurement. Whereas for small systems the indices can be provided via manual input, performing the selections within larger systems such as protein complexes is best performed using external programs (e.g. VMD). The indices can be imported as plain text files containing one index per line. Very closely related to the DCDRMSD2D program is the DCDMultiTraj program, which also calculates RMSD values in a two dimensional manner. In contrast to the DCDRMSD2D program, the comparisons are not performed on the frames of a single trajectory. As indicated by its name, the DCDMultiTraj program compares each single geometry of one dcd file mutually with one of a second dcd file. The index numbers along the first trajectory are used as abscissa while the ordinate gives those along the second trajectory. As the two trajectories do not need to have the same number of frames, the resulting graph is not necessarily square. Nevertheless, the two trajectories have to been derived from the same system, e.g. the number of atoms have to be identical.

5.2 Modification of Molecular Geometries

One of the most frequently occurring task in the preparation of the calculations carried out within this work is the modification of the molecular geometry. Especially for the computations of reaction profiles, incremental

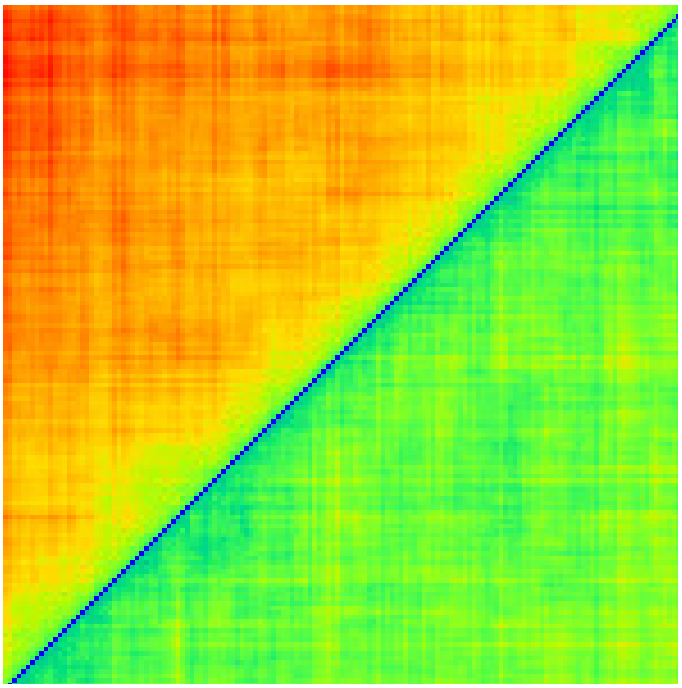


Figure 5.1: Example of a two-dimensional RMSD plot generated with the DCDRMSD2D program. The values in the lower right triangle represent RMSD values calculated between the atoms used to align the geometries (backbone N,Cα,C,O) whereas the ones in the upper left triangle provide an additional measure on a second selection of atoms (protein heavy atoms). The color range is 0.0Å (blue) to 2.0Å (red).

changes of bond distances, bond angles or torsion angles are required. To facilitate these modifications, a suit of programs was developed. All programs for a specific type of coordinate share a common scheme whereas different subroutines are used for reading and writing the appropriate file formats. Modifications are available for the xyz, pdb and the Turbomole[42] coord file format. Furthermore, measuring internal coordinates in AIMpac wavefunction (wfn) files is also supported while modifying them is not.

5.2.1 Modification of Bond Distances (...BOND)

Bond distance are the most trivial type of internal coordinates and are the easiest to be modified. A chemical bond, or any linear path between two atoms A and B, can be represented as vector $\overrightarrow{a} = \overrightarrow{AB}$ with its length giving the interatomic distance between the two nuclei. Conversely, scaling the vector with its foorpoint at A, so that its length equals a chosen value, determines the position of atom B after the modification. The corresponding programs are designed to be used via a single command line input in the form:

```
...BOND <infile> <atomA> <atomB> <value> <outfile>
```

The `<infile>` argument specifies the file containing the inital set of coordinates. `<atomA>` and `<atomB>` are the indices used to define the endpoints of the bond vector. While these three arguments are mandatory, the other two are optional. If the later are omitted, the programs perform only a measurement on the selected distance. If a specific value, `<value>`, is provided and an output filename, `<outfile>`, is specified, the chosen bond distance will be adjusted to the provided value and the results will be written to the specified file. Whereas the value provided by the user is assumed to be given in Å, the I/O subroutines assure the conversion into the suitable length units depending on the file format. If the `<infile>` and `<outfile>` are identical, the input file will be updated with the modified geometry without further notification. After reading the initial geometry from the input file, using the corresponding subroutines, the bond vector $\overrightarrow{a}$ is constructed as difference between the position vectors of A and B. The vector is then scaled to the desired length, yielding $\overrightarrow{a}'$. Finally, the new position for atom B is determined as $\overrightarrow{B}' = \overrightarrow{A} + \overrightarrow{a}'$. The output is then written using the corresponding subroutine.

5.2.2 Modification of Bond Angles (...ANGLE)

For the modification of bond angles, the corresponding angle has to be defined by three atoms, A, B and C, such that B is the pivot and C will be the atom to be moved during the adjustment. For the modifications, two bond

vectors with footpoint at B are defined, $\overrightarrow{a} = \overrightarrow{A} - \overrightarrow{B}$ and $\overrightarrow{b} = \overrightarrow{C} - \overrightarrow{B}$. The bond angle can be represented as $A \leftarrow B \rightarrow C$. The command line to invoke the program has the form:

```
...ANGLE <infile> <atomA> <atomB> <atomC> <value> <outfile>
```

The meaning of the command line arguments is identical to that for the ...BOND programs except for the `<atomC>` argument, which specifies the index of the third atom. The modification of the bond angle is preceded by rotations of the initial geometry around the coordinate axis, such that the vector $\overrightarrow{a}$ is parallel to the x-axis and that the z-component of the vector b vanishes. This reduces the problem of modifying the angle between two vectors in three dimensional space to the adjustment of the angle between the x-axis and vector $\overrightarrow{b}$ in only two dimensions, depicted in Figure 5.2.

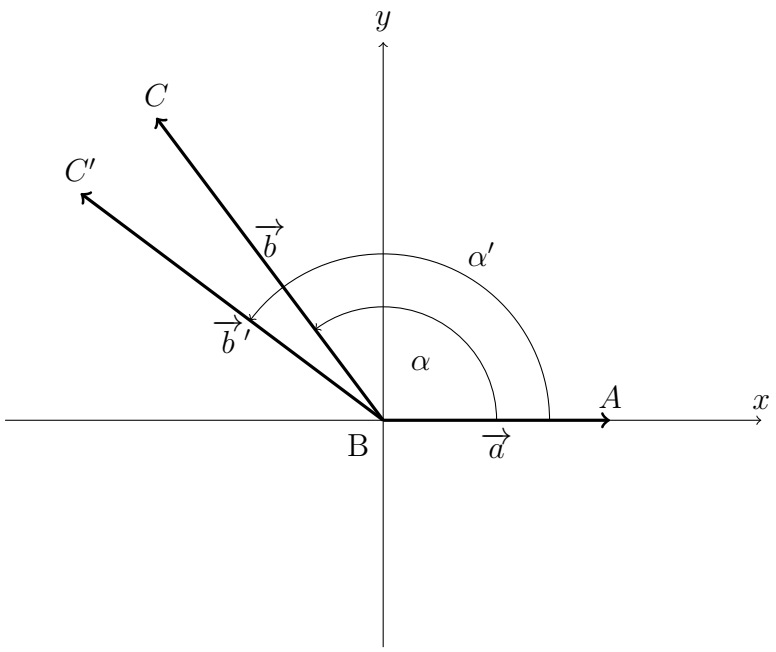


Figure 5.2: Scheme of modifying bond angles in the plane spanned by the corresponding bond vectors.

For a specified value of the bond angle α', the updated coordinates of atom C can be calculated by trigonometry as $C'_x = |\overrightarrow{b}|cos(\alpha')$ and $C'_y = |\overrightarrow{b}|sin(\alpha')$. Finally, the initial rotations are reversed and the updated geometry is written to the specified output file using the appropriate I/O routines.

5.2.3 Modification of Torsion Angles (...DIHED)

For the last of the internal coordinates, torsion angles, four atoms are needed in the definition, namely A, B, C and D. The torsion angle can then be written as $A \leftarrow B \rightarrow C \rightarrow D$ with the three vectors $\overrightarrow{a} = \overrightarrow{BA}$, $\overrightarrow{b} = \overrightarrow{BC}$ and $\overrightarrow{c} = \overrightarrow{CD}$ (see Figure 5.3, left). Similarly to the programs described above, the modification of the torsion angle is accomplished by moving only atom D. The program is called via:

```
...DIHED <infile> <atomA> <atomB> <atomC> <atomD> <value> <outfile>
```

the meaning of the arguments again resembles that of the ...BOND and ...ANGLE programs with `<atomD>` specifying the fourth atom needed to define the torsion angle. To accomplish the measurement of the torsion angle of the initial geometry, two additional vectors are constructed using the cross product $\overrightarrow{b} \times \overrightarrow{a} = \overrightarrow{n_1}$ and $\overrightarrow{b} \times \overrightarrow{c} = \overrightarrow{n_2}$. Thus the torsion angle can be determined via the scalar product of $\overrightarrow{n_1}$ and $\overrightarrow{n_2}$ to be $\theta = acos\left(\frac{\overrightarrow{n_1} \cdot \overrightarrow{n_2}}{|\overrightarrow{n_1}| \cdot |\overrightarrow{n_2}|}\right)$

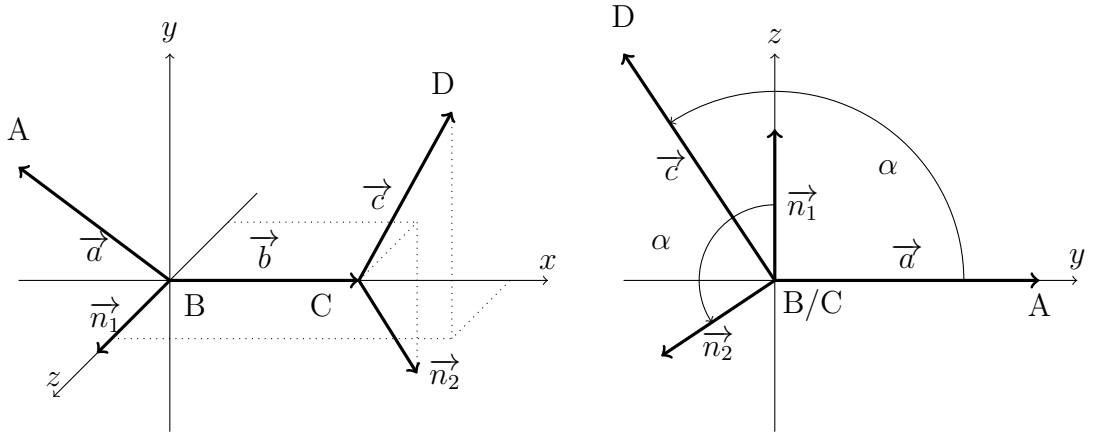


Figure 5.3: Scheme of vectors used to modify dihedral angles before the rotation.

Although depicted with the $\overrightarrow{b}$ vector being parallel to the x-axis in Figure5.3, this is no prerequisite for the determination of the actual dihedral angle in the provided structure. However, if a modification of the dihedral angle is requested, the calculations are facilitated by preceeding rotations of the system, such that the orientation corresponds to that shown in Figure 5.3 (left side). With the vector $\overrightarrow{b}$ aligned to the x-axis and the vector $\overrightarrow{a}$ being part of the xy-plane, the dihedral angle reduces to that between the y-axis and the projection of $\overrightarrow{c}$ onto the yz-plane or simply the angle between

the z-axis and $\overrightarrow{n_2}$. A modification of the dihedral angle thus simplifies to a rotation of $\overrightarrow{c_{yz}}$ while its x-component remains unchanged. Once the dihedral angle is adjusted, the preceeding rotations of the system are reversed and the modified coordinates of atom D are printed to the output file.

5.3 Generation of Molecular Clusters

The investigations on the electron density of a single molecule within its crystal environment have been carried out identically to those performed on the complexes with the target enzyme. Those QM/MM calculations require the system to be specified according to the specifications of the Charmm force field and, preferably, the geometry being provided as pdb (protein database) file. Each individual molecule within the crystal should be named according to the topological definitions of the force field, each one being treated as separate residue. Unfortunately, despite the large number of programs being able to generate clusters of arbitrary size based on a cif (crystallographic information file), none of them could fully satisfy the needs described above. This led to the development of the `Decifer` program. The `Decifer` program uses the data from a given cif, yet due to the flexibility of the file format specifications, reading directly from the cif is not supported. Instead, the program uses two separate input files which can readily be derived from the underlying cif. Examples of two such inputs are provided in Figures 5.4 and 5.5. The definition of the unit cell within the first of these files follows the general scheme, in which the side a is oriented parallel to the x-axis of the cartesian coordinate system and b is part of the xy-plane. The definitions of the unit cell are followed by the number of atoms and the actual geometry of the molecule. There is one line for each atom, which must contain an atom label, the element symbol and the fractional coordinates of that atom. There is no fixed format for these entries except that the fractional coordinates have to be given as pure floating point numbers. If the experimental uncertainties are given, denoted by parenthesis, they have to be removed. The assigned atom labels must correspond to the ones used in the topological definition of the molecule within the Charmm force field. Additional information at the end of the line might be present but will be discarded by the `Decifer` program.

The second input file (Figure 5.5) contains the symmetry operations of the crystal. It starts with the number of symmetry operations, followed by each single operation given as a comment line, a 3×3 rotation matrix and a translation vector, both given in fractional coordinates. The first operation is usually the identity, represented by the 3×3 unit matrix and the zero vector. The other operations can be derived from the symmetry operations provided in the cif as depicted in Figure 5.5.

```
_cell_length_a                    7.9295
_cell_length_b                    11.306
_cell_length_c                    14.113
_cell_angle_alpha                 90.00
_cell_angle_beta                  99.94
_cell_angle_gamma                 90.00
29
S1 S 0.02091 0.390560 0.147255 Uani 1.000000000
S2 S -0.49466 0.520331 0.160684 Uani 1.000000000
O3 O 0.07478 0.49835 0.10628 Uani 1.000000000
...
```

Figure 5.4: Example of the structure input file for the program. The data represents the unit cell definitions and the first atoms of TS793 taken from the crystal structure.

```
4
#1st operation      x, y, z
1. 0. 0.
0. 1. 0.
0. 0. 1.
0. 0. 0.            no translation
...
#4th operation      x+1/2, -y+1/2, z+1/2
1. 0. 0.
0. -1. 0.
0. 0. 1.
0.5 0.5 0.5         +1/2 unit cells in each direction
```

Figure 5.5: Example of the symmetry input file used for the Decifer program. Each symmetry operation if given via its 3×3 rotation matrix and a translation vector.

For the generation of the output structure, the `Decifer` program used cascaded loops for all symmetry operations and spatial dimensions, generating separately each single monomer within the crystal. The coordinates of each monomer are based on the fractional coordinates taken from the input. In a first step, the symmetry operation is applied to the molecule by multiplication of the $3x3$ symmetry matrix with the fractional coordinates followed by the addition of the translation vector. The so generated molecule is then translated along each of the crystallographic axis by adding the corresponding integer number to the atomic coordinates. Finally, the modified fractional coordinates are converted to cartesian ones by multiplication of a transformation matrix, comprising the cartesian representation of the unit cell vectors. Within the output file, the cartesian coordinates are augmented with the topological data of the atoms as well as a sequential and continuous sequence of atom and residue index numbers.

5.4 Potential Energy Surfaces

For the investigations of chemical reactions, a great interest lies in the relative energies of the reactant and product structures as well as that of any intermediates along the path between them. Whereas this dependence is usually visualized in simple two-dimensional energy profiles, one being the *reaction-coordinate*, the other one being the (relative) energy, its physical origin is much more complex. The reaction coordinate, although being used as one-dimensional coordinate, covers all structural changes along the reaction path, thus comprising changes in every individual internal coordinate of the system. Nevertheless, not all these changes are equally important in the description of the reaction. As most of them do hardly change during the reaction, they do not contribute to the diversity of the structures. Consequently, only those internal coordinates which exhibit significant changes have to be considered. For common S_N2 reactions at carbon center, the reaction coordinate can sufficiently accurately be approximated by the distance between the carbon atom and the attacking nucleophile. There are, however, significant changes in other coordinates such as bond angles or the distance between the carbon atom and the leaving group. Nevertheless, these changes are coupled to the chosen reaction coordinate and can thus be neglected. Slightly more complicated are β-elimination reactions of E1, E2 or E1cB type, which require at least a two-dimensional representation, with respect to spatial coordinates, to allow a distinction between the three. As leading coordinates, chemists usually chose the distance between the leaving group and the adjacent carbon atom as first, and the distance between the β-carbon atom and the removed hydrogen atoms as second coordinate. This allows to determine, whether the leaving group is ejected before (E1) or after

(E1cB) the abstraction of the hydrogen atom, or if both processes happen simultaneously (E2). Using a two-dimensional representation of the corresponding energies in form of contour lines is the basis of the famous More O'FerrallJencks plots tough in organic chemistry lectures. From a technical perspective, potential energy surfaces can be generated by starting from a given molecular structure and adjusting the chosen coordinates in small increments until the other end of the reaction path is reached. At each increment, the geometry is adjusted and optimized by relaxing all coordinates but the ones being modified. As this procedure requires repetitive tasks, a suite of programs was developed to facilitate the generation of potential energy surfaces for quantum mechanical calculations with the Turbomole package as well as combined QM/MM calculations within ChemShell using either the Charmm or Amber force field. Depending on the package used for the actual calculations, the PES program suite carries different prefixes. `QM...` indicates purely quantum mechanical calculations with the Turbomole program. Combined quantum mechanical / molecular mechanical calculations using the Charmm force field are denoted `QMMM...`, whereas those employing the AMBER force field are named accordingly `AMBER...`. To perform the calculations needed to generate a PES, each coordinate is decomposed into equidistant grid points, resulting in a two dimensional grid of cells. Each of them corresponds to an individual calculation. The data structure used to contain these calculations consists of folders and subfolders named according to the corresponding coordinate on the grid. The coordinates are translated into 3 digit wide integer numbers. Bond distances are given in $\frac{1}{100}$th Å while bond and dihedral angles are given in degrees. Assuming a calculation is performed for a bond distance of 1.75Å along the first and 85 degrees along the second dimension, the corresponding folder would be named `../175/085/`. To manage the individual cells of the potential energy surface and the tasks to be performed for each of them, the program package works similar to a cellular automaton. These programs work iteratively while in each cycle, the new state of each cell solely depends on its old state and that of all its neighbors. Differences are found in the dimension of the grid and the rules, which are used to determine the status of each cell at each cycle. Famous examples of cellular automata are Conway's Game of Life, the simulated planet WATOR[202] or the Nagel-Schreckenberg model[203] used to model the emergencse of traffic jams. The potential energy surface program package described here however differs significantly from these examples in some major functions. Whereas all examples are designed as simulations that run continuously at constant increments of time, the described program package performs only a single cycle each time it is invoked by the user. Furthermore, as the major goal was not to construct a simple simulation model but to facilitate the preparation and execution of calculations, each individual cell can have far more diverse states than the usual on/off discrimination used in

most of the cellular automata. The possible states as well as the algorithms to determine the correct ones are described in the following sections.

5.4.1 General Features

With the experience from previous attempts to automate the generation of potential energy surfaces, the current package should also be as modular as possible, providing a central part for the automation as well as several status programs for displaying the current state of each cell (... converged), extracting the obtained energies (... energies) such that they can be visualized with external programs and collecting the optimized geometries (... structures), providing the possibility for a visual inspection of the results. The prefix of these programs depends on the applied methods being pure quantum mechanics (TM...), QM/MM using Charmm/Turbomole (QMMM...) or Amber/Turbomole (AMBER...). For correct functionality, all these programs need information about the dimensions of the potential energy surface as well as their boundaries and increments. Hence, a central file is used ("settings.PES"), in which this information is provided as user readable plain text. An example for the settings file is given below.

The first section is used to define the type of internal coordinate being used for either dimension. Possible selections are bond distances, bond angles of dihedral angles. The alternative option "n" may be used if a single dimension is sufficient to describe the reaction profile. In this case, the molecular geometry will only be modified according to the other dimension. With the type of the dimension known, the corresponding atomic indices have to be provided in the next section. Due to technical reasons, atoms A to D are always assigned to the first dimension while E to H define the coordinate for the second one. As described in the file, bond distances require only two atomic indices and bond angles three. Specifications of more than those atomic indices required to properly define the selected type of coordinate is possible, but those not needed will be discarded. If fewer indices as those required are given, the PES programs will abort with a corresponding message. In some special cases, such as modeling a proton transfer via the two distances between the hydrogen atom and the donor and acceptor, respectively, the indices used to define the two coordinates might overlap. If so, the coordinates should be defined in such a way, that at least their last index numbers do not occur in the definition of the other coordinate. This will avoi inaccuracies originating from the way, the modification of coordinates is carried out. In the following section, the boundaries and the increments along both dimensions have to be defined. Distances should be given in $^1/_{100}$Å, while angles and dihedrals are to be specified in degrees. Obviously, the Amin and Bmin values must be lower than their Amax and Bmax counterparts. Furthermore, the increments Astep and Bstep should be given such

```
Definition of the constrained internal coordinates:
DIMA=b
DIMB=b
b:    bond
a:    angle
d:    dihedral
n:    none

Definitions of internal coordinates by atom indices
 bonds:          AB   or EF
 angles:         ABC  or EFG
 dihedrals:      ABCD or EFGH
atomA=2281
atomB= 347
atomC=
atomD=
atomE=2281
atomF=2280
atomG=
atomH=

Definition of min/max values for chosen dimensions
 lengths are given in angstrom units (120 = 1.20A)
 angles/dihedrals are specified in degrees
Amin=110
Amax=220
Astep=10
Bmin= 80
Bmax=210
Bstep=10

Here is the label that will be used for the queuing system:
LABEL=CCm_

OPTIMIZATION WALLTIME=323:00:00
CPUNUM=2
```

Figure 5.6: Example of a "settings.PES" file using two bonds (H_{2281}-S_{347} and H_{2281}-N_{2280}) as coordinates to model the proton (H_{2281}) transfer between a cystein (S_{347}) and a histidine (N_{2280}) residue..

that a reasonable stepsize between the boundaries is provided. The cell with the lowest coordinates will be at Amin/Bmin. With these parameters, the generation of the potential energy surface could technically be started. As the package however interfaces the queuing system of the cluster, further entries my be specified. The given label can be used to locate submitted calculations in the list of a users jobs currently being processed by the cluster. The complete name of the job uses the label as prefix completed by the two three digit wide integers representing the coordinates of the current cell. To most efficiently use the available resources, the maximum walltime can be specified for each job. Parallelization is available via CPUNUM, whereas a value of 1 disables any parallelization.

Besides the general settings file, a folder named "templates" should be present, in which files are to be provided, which a used unmodified at each point of the surface. Whenever a calculation is being prepared, those files will be copied to the current cell. Which files need to be provided depends on the type of calculations for which three options are described below.

The last component needed for the generation of a complete potential energy surface is a starting geometry. This should be preferably a structure already optimized at the same level of theory that will be used for all successive calculations. This preliminary calculation should be placed in a subfolder that corresponds best with the optimized geometry. Depending on the programs used for the actual calculations, different files need to be present as describe below.

5.4.2 Quantum Mechanical PESes

The most simple type of calculations from a technical point for view are quantum mechanical calculations performed with the Turbomole[42] package. The corresponding package for the PES generation provides a two-leveled system of optimizations and successive single-point calculations for all cells of the surface. This allows to optimize the molecular geometry at a computationally less demanding method while calculating the energies at a second, more accurate level assures more realistic energies, especially concerning those of transition states. To start a potential energy surface, one of the cells needs to contain the initial geometry stored in the Turbomole coordinates file format with the file being name "coord-final". Furthermore, a file name "GEO_OPT_CONVERGED" needs to be present, indicating a successful optimization. To prepare each individual calculation, templates for the "define" and "cosmoprep" tools have to be provided in the "templates" subfolder. The filenames should be "optdefine.inp", "spdefine.inp" and "cosmo.inp". If the latter is missing, the package will prepare calculation in vacuum. If provided, the file "cosmo.inp" will be used to parameterize the COSMO before submitting the calculation to the queuing system.

Preparing the calculations is performed by invoking the "auto.PES" script, which should either be place in the main directory of the potential energy surface of in a folder being defined in the users "PATH" variable. The script then uses the parameters provided in the file "settings.PES" to perform a first scan of the folder structure. Wherever a converged calculations is encountered, indicated by the presence of the file "GEO_OPT_CONVERGED", all folders of cells adjacent to the current one are created if they not already exist. After that, the script performs a second scan over all cells in which the status of each cell is determined and the corresponding actions are performed. A Scheme of this algorithm is given in Figure 5.7 and 5.8. Depending on the state of each cell, the required action might be to skip the cell, to prepare an optimization or to prepare a single-point calculation. Whenever an optimization is to be prepared, all existing files in the corresponding subfolder except for the file "source" will be deleted. Then, a suitable starting geometry is seeked for in the adjacent folders. As optimization might be prepared from up to eight possible geometries, the file source keeps track of any geometry already used. If a geometry was already used before and is listed in the file, it will no longer be considered as potential starting geometry. If however a starting geometry is found, it will be copied to the current folder, modified according to the actual coordinates and the calculation is prepared using "define" and, if needed, "cosmoprep". After that, a queuing script is generated using the label, walltime and number of CPUs to be used from the settings file. Finally the calculation is submitted to the queuing system and the script advances to the next cell. Preparing a single-point calculation is performed as soon as the preceeding geometry optimization has succeeded. If so, a subfolder named "SP" is created and the optimized geometry is copied there. The calculation is then prepared analogously to that before but using the appropriate "spdefine.inp" template rather than that for an optimization. The template for "cosmoprep" is identical for both types of calculations. Once prepared, a queuing script is generated for the single-point calculation and its being submitted to the computing cluster. To keep track of optimizations and single-point calculations, the assigned jobnames carry the suffix "o" and "s", respectively.

The first check performed for each cell is to determine whether the current optimization has already finished successfully. As Turbomole generates the file "GEO_OPT_CONVERGED" in this case, the presence of this file is tested. If the file is found, the script branches to the preparation of single-point calculations described in Figure 5.8, otherwise the optimization pathway will be followed. Potential errors may occur within Turbomole, in which case the file "GEO_OPT_FAILED" will be be generated. If this file is encountered, an optimization will be prepared using another starting geometry as that of the faulty calculation. In case this file can not be found, the next one inquired is the file "source". As long as this file can not be found,

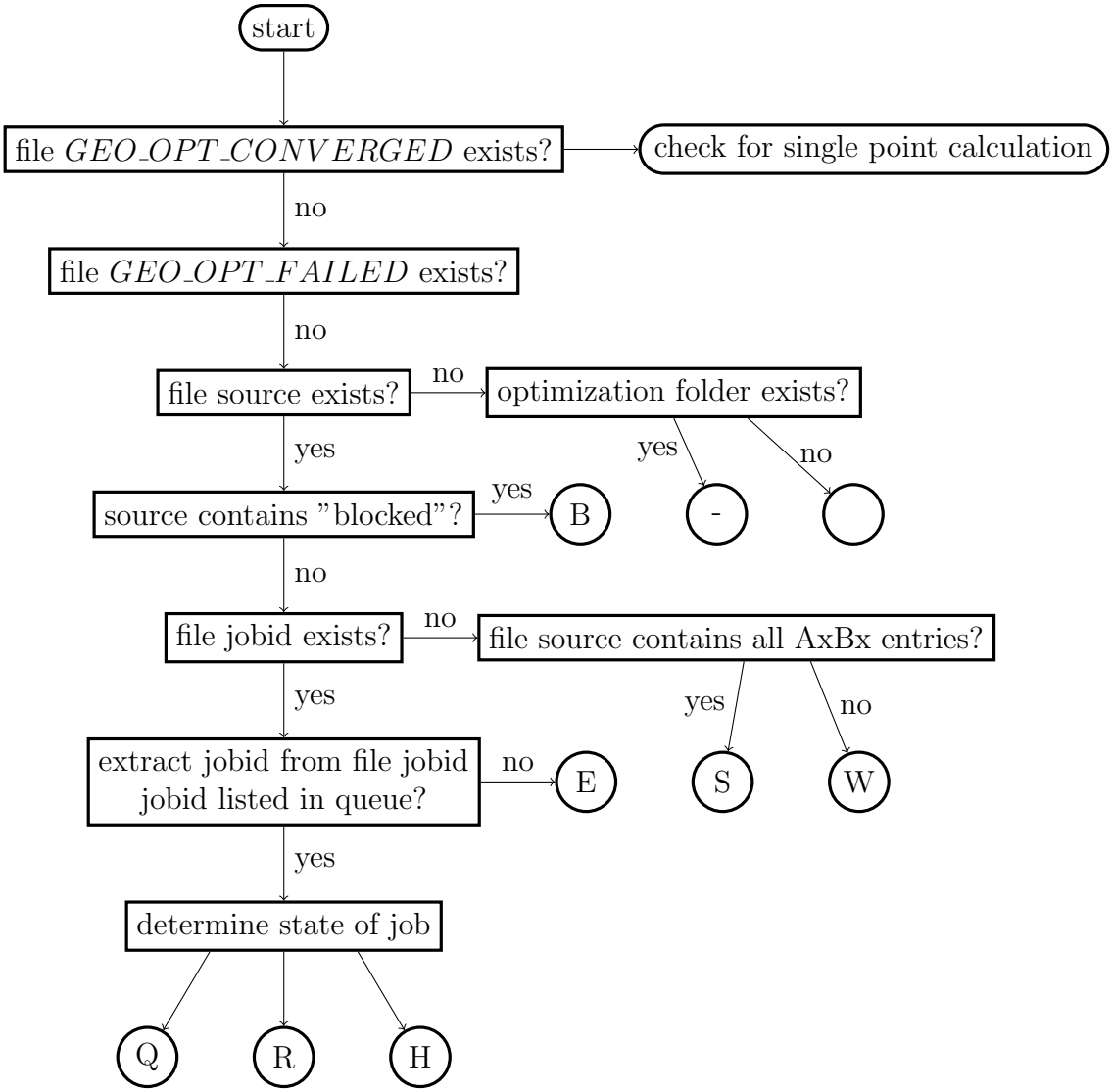


Figure 5.7: Program flowchart indicating the determination of tasks to be performed to optimize the geometries of a Turbomole potential energy surface.

there are only two possibilities left. Either the subfolder corresponding to this cell does not yet exist, then the cell is being skipped. If the folder exists, an attempt will be made to prepare the first optimization as described before. As a visual aid during execution, the script writes characters to the screen indicating the determined status for each cell. A non-existing subfolder will result in a blank. While the script attempts an optimization once the subfolder exists, the "TMconverged" program simply writes "-" for this cell. If the file "source" can be found, the cell is either blocked from being calculated or a previous optimization was already started. The feature of blocking a cell was introduced to reduce the computational effort by avoiding to perform calculations on regions, which are not of interest to the user. To block a cell, the file "source" has to exist and it must contain the string "blocked". Th cell is being skipped and the display shows "B". If the cell was not lbocked, the programs have to determine if the previously prepared job is currently being processed. This might be the case if the file "jobid" can be found. "jobid" will be created upon submitting a job to the queueing system. It contains the assigned job identification number, so that the job can be easily identified later on. If the "jobid" file could not be found, this usually indicates that no suitable starting geometry for the current cell was found. Therefore, if the file "source" contains the identifiers of all surrounding cells (AxBx entries), there can simply be no suitable geometry whereas if not all these entries are present, the script is just waiting for the next potential one to become available. Correspondingly the status symbols are "S" if no suitable source exists and "W" indicating that the program is waiting for the next one. If however the jobid can be found in the corresponding file, the queuing system is inquired for the actual status of that job. If the queuing system does not show the identifier, an error has occured, which the program suite does not automatically handle. Displaying "E" requires the user to investigate the origin of that error manually. If the queuing system however does list the job, the current state might be "R"unning, "Q"ueued or "H"old. In any of the later cases, no further actions are performed for this cell.

Once the geometry optimization was successful, the preparation of single-point calculations becomes possible. If however the file "spdefine.inp" does not exist in the "templates" folder, no such calculations will be prepared. Thus also no "SP" subfolders will be created and the status of the cell remains "C"onverged. With the calculation of single-point energies being requested, "SP" subfolders will be generated and the corresponding calculations will be started as soon as the optimized geometry is available. A finished single-point calculation is indicated by the presence of the file "dscf.out". If this file contains the entry "all done", the calculation succeeded and the corresponding energy will be used by the TMenergies program. If the entry is missing, the calculation experienced technical issues to be checked by the user. The latter

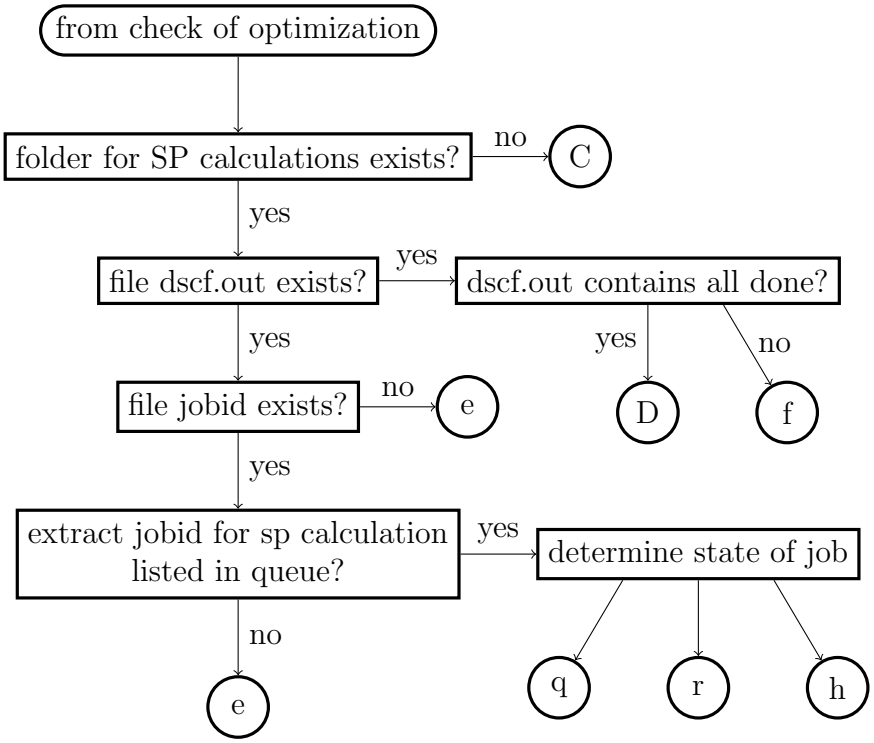


Figure 5.8: Program flowchart indicating the determination of tasks to be performed for the single point calculations upon the geometries obtained during the optimizations.

is indicated by "f". In very much the same way as described for the optimizations, the job identifier is used to determine the status of the job within the queuing system. The letters indicating the possible states are identical to those of the optimizations, yet single-point calculations are labeled using lower case letters to distinguish them from optimizations, marked with upper case letters. The labels are also used in the TMconverged, whereas this program additionally provides a numerical summary of all jobs of a PES calculation based on their actual status.

The TMenergies program also uses the same scheme described above although only successful calculations are being considered. All energies from finished calculations are extracted. Depending on the users choice, the energies are converted from atomic units to $\frac{kJ}{mol}$, $\frac{kcal}{mol}$ or eV. Then all energies are normalized to the most negative one before being printed to the files "energies.txt" or "spenergies.txt" for optimizations and single-point calculations, respectively. The values are written in ASCII encoding in the form of first coordinate, second coordinate and energy value. Thus they can readily be displayed in gnuplot or other visualization tools.

TMstructures can be used to extract the final geometries from each cell. The converged geometries are converted to the standard xyz file format and a file corresponding to each optimized geometry is created in the folder "structures". These files can be accessed with any molecular file viewer and provide a convenient means of verifying the correlation of the structures along the surface as well as to monitor the geometric changes along the reaction path.

5.4.3 PESes with Turbomole and Charmm

Based on the programs developed for pure qantum mechanical calculatons with Turbomole, a similar package was developed for QM/MM calculations with ChemShell using Turbomole for the QM- and Charmm for the MM-subsystem. Due to issues with the combination of optimizations and single-point calculations using different functionals for the QM-subsystem, only the preparation of QM/MM optimizations was implemented. Working in principal identically to the previously described package, there are some distinct differences between the two packages. Most obviousl, the files that need to be provided in the folder "templates" are different. For the QM/MM calculations, these files are the topological definitions and parameter set for the force field stored in the files "top.rtf" and "par.prm". Furthermore, the explicit topology of the current system, the protein structure file "qmmm.psf", needs to be provided here. Also the files "act", defining the active region of the calculation, and "oxys" specifying the water molecules need to be provided. Also the files "save_qmmm.chm" with further definition of the system and "qmmm_opt.chm" with the instructions for ChemShell are needed.

As the QM/MM calculations depend on different files than pure quantum

mechanical calculations, also the scheme to determine the cell status needed adjustments. They are sketched in Figure 5.9.

The check for a successful optimization in the case of QM/MM calculations depends on the output file "qmmm_opt.out", which should contain the keyword "Optimization finished successfully". The alternative phrase "Converged!" is used in combination with the AMBER force field described below. The remaining checks are comparable to those of the pure QM calculations as the corresponding files are created by the PES program package rather than by the programs that actually perform the calculations. The preparation of QM/MM calculations however is much more easy than that of pure QM calculations. It only requires the modification of a suitable starting geometry, using the PDB... programs rather than their TM... couterparts, and the files of the "templates" folder to be copied to the actual cell. Any further preparation is already included in the template files or will be performed within ChemShell once the cluster system has started the optimization.

As for the previous package, specialized programs for the analysis are available. QMMMconverged provides a fast display of the current status of all cells without further influencing the calculations. QMMMenergies performs a similar extracting of the minimized energies along the surface. And finally, the QMMMstructures program provides a means to collect all optimized geometries in the "pdbs" subfolder for further analysis.

5.4.4 PESes with Turbomole and Amber

A second package for QM/MM calculations was developed for calculations which use the AMBER force field rather than Charmm for the MM-subsystem. However, as both types of QM/MM calculations are carried out with ChemShell, the general structure of the automation scripts is almost identical. The scheme to determine the cell status indeed is identical except for the keyword "Converged!" with which a successful optimization is indicated. Furthermore, QM/MM calculations with AMBER require a different set of template files, namely "act", "LIG.frcmod", "LIG.lib", "qmmm_opt.chm"and "qmmm.prmtop". Compared to the calculations with Charmm, an additional preparation step is required for each optimization. Due to some issues with the "tleap" program of AMBER in connection with the ChemShell interfaces, the preparation of the "qmmm.rst" file was included into the preparation of each individual calculation by the tool "AMBERRST". As the rst file requires the cartesian coordinates of the system, the file can not be provided as global template but must be generated separately for each calculation. Therefore, the cartesian coordinates of the system are read from the pdb file and a corresponding rst file is being created. As described before, a set of programs is available to extract energies (AMBERenergies) and structures (AMBERstructures) as well as to display the status of the current

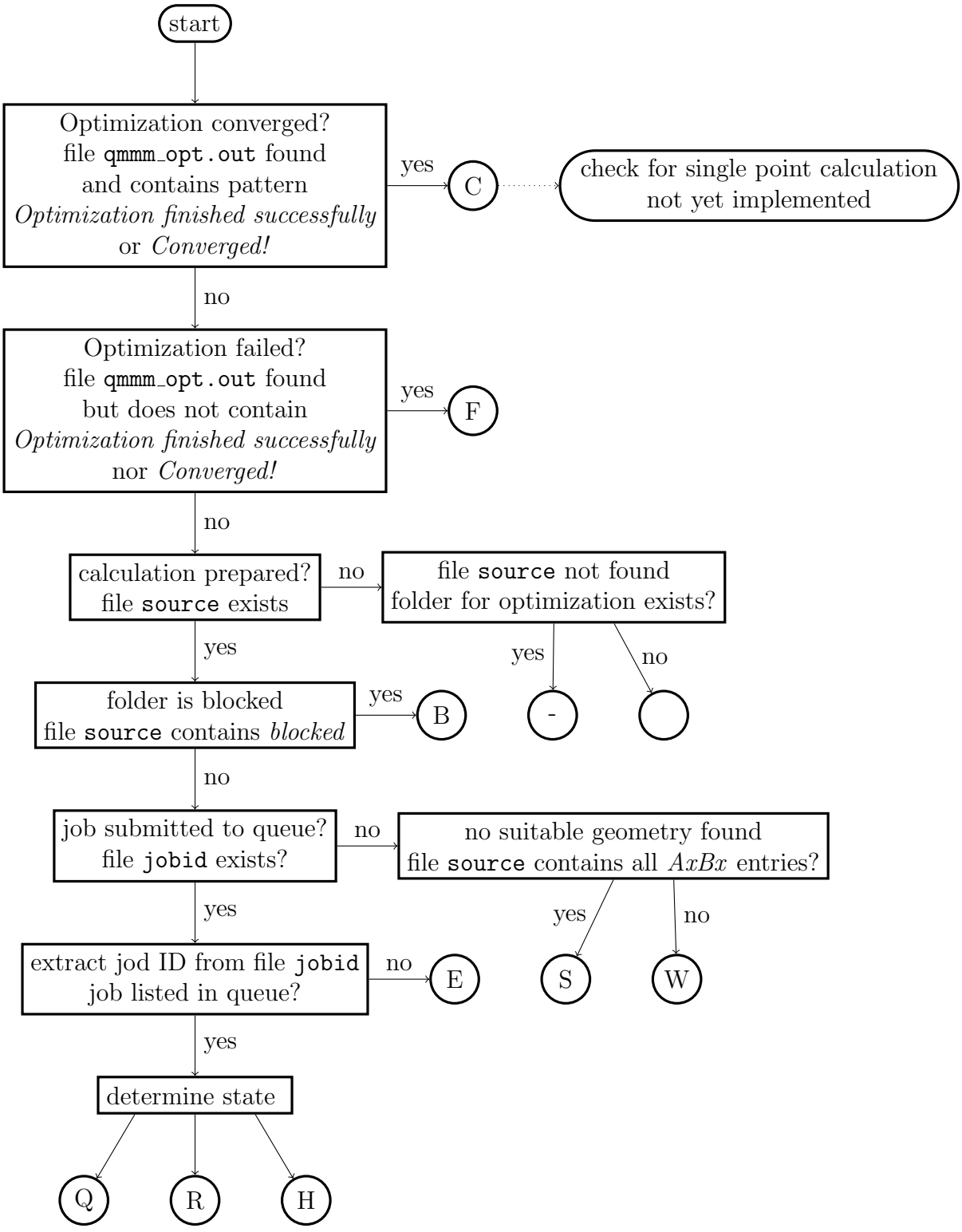


Figure 5.9: Program flowchart indicating the determination of tasks to be performed to optimize the geometries along the potential energy surface computed in the framework of QM/MM.

PES (AMBERconverged).

6 Summary

In the first part of this work, a combination of theoretical methods for the rational design of covalent inhibitor is presented. Starting from the crystal structure of the covalent complex of a lead compound, quantum mechanical and QM/MM calculations were used to derive the exact geometry of the preceeding non-covalent enzyme inhibitor complex. The geometry of the latter mainly determines the reactivity of the inhibitor against its target enzyme concerning the formation of the covalent bond towards an active site residue. Therefore, this geometry was used as starting point for the optimization of the substitution pattern of the inhibitor such as to increase its binding affinity without loosing its ability to covalently bind to the target protein. The optimization of the chemical structure was supported by using docking procedures, which are best suited to estimate binding affinities that arise from the introduced changes. A screening of the novel substitution patterns resulted in a first generation of model compounds which were further tested for their reactivity against the target. Dynamic simulations on the novel compounds revealed that the orientation that compounds adopt within the active site are such that a covalent interaction with the enzyme is no longer possible. Hence, the chemical structure was further modified, including not only changes in the substituents but also within the core of the molecule. Docking experiments have been conducted to assure sufficiently high binding affinities and to obtain the most favored binding poses. Those have then again been used for dynamic simulations which resulted in structures, for which the bond formation process appeared feasible. A final series of QM/MM calculations considering various protonation states was computed to estimate the reaction energies for the covalent attachment of the inhibitor to the enzyme. The theoretical results indicate a reasonable high inhibition potency of the novel compounds.

The second part concentrates on the environmental influences on the electron density of an inhibitor molecule. Therefore, a vinylsulfone-based model compound was selected for which an experimental crystal structure for the pure compound as well as a theoretically determined enzyme-inhibitor complex have been available. To provide reference data for the larger systems, the conformational space of the isolated molecule was screened for favorable geometries which were later compared to those within the crystal and protein surrounding. The geometry of the crystal structure could readily be taken from the experimental data whereas calculations on the protein

complex revealed four potential non-covalent complexes exhibiting different arrangements of the molecule within the active site of the protein as well as two possible protonation states of the catalytic dyad. Hence, all four protein complexes have been compared to the crystal structure of the molecule as well as against the more favorable geometries of the isolated molecule being determined within vacuum or aqueous surrounding. Whereas the molecule itself was found to adopt comparable geometries within all investigated environments, the interactions pattern between the crystal surrounding and the protein differed largely from each other. The favorable formation of dimers within the crystal has a strong stabilizing effect and explains the extraordinarily good quality of the crystal. Within the protein however, repulsive forces have been found between the protein and the inhibitor. The origin of the repulsion could be traced back to effect of on of the substituents to the vinyl scaffold. The difference in the chemical structure in comparison to a well known inhibitor might also explain the experimentally found loss of activity for the model compound in comparison to K11777.

6.1 Zusammenfassung

Im ersten Teil dieser Arbeit wird eine Kombination theoretischer Methoden für die strukturbasierte Entwicklung neuer Wirkstoffe präsentiert. Ausgehend von der Kristallstruktur eines kovalenten Komplexes einer Modellverbindung mit dem Zielprotein wurde mit Hilfe von quantenmechanischen und QM/MM Rechnungen die genaue Geometrie des vorausgehenden nicht-kovalenten Komplexes betimmt. Letztere ist der bestimmende Faktor für die Reaktivität des Inhibitors gegenüber der katalytisch aktiven Aminosäure und damit für die Ausbildung einer kovalenten Bindung. Aus diesem Grund wurde diese Geometrie auch für die Optimierung der Substitutionsmusters des Ihnibitors verwendet, um dessen Affinität zum Zielenzyme zu verbessern ohne dass dieser seine Fähigkeit kovalent an das aktive Zentrum zu binden verliert. Die Optimierung des Substitutionsmuster wurde doch Methode des Molekularen Dockings unterstützt, das diese optimal dazu geeignet sind, Bindungsaffinitäten vorherzusagen, die durch eine Modifikation der chemischen Struktur entstehen. Eine Auswahl der besten Strukturen wurde anschließend verwendet, um zu überprüfen, ob die veränderten Moleküle noch genügen Reaktivität gegenüber dem Zielprotein aufweisen. Moleküldynamik Simulationen der neuen Verbindungen haben jedoch gezeigt, dass die veränderten Verbindungen nur so and das Protein binden, dass die Bilung eine kovalenten Bindung zum Enzym nicht mehr möglich ist. Daher wurden in einem weiteren Schritt die Modellverbindungen weiter modifiziert. Neben Änderungen im Substitutionsmuster wurde auch die chemische Struktur im Kern verändert. Die Bindungsaffinitäten

wurde wieder mittels Docking überprüft. Für die besten Bindungsposen wurden wieder Simulationen zur Moleküldynamik durchgeführt, wobei diesmal die Ausbildung einer kovalenten Bindung zum Enzyme möglich erscheint. In einer abschließenden Serie von QM/MM Rechnungen unter Berücksichtigung verschiedener Protonierungszustände des Inhibitors und des Proteins konnten Reaktionspfade und zugehörige Reaktionsenergien bestimmt werden. Die Ergebnisse lassen darauf schließen, dass eines der neu entwickelten Moleküle sowohl eine stark verbesserte Bindungsaffinität wie auch die Möglichkeit der kovalenten Bindung an Enzyme aufweist.

Der zweite Teil der Arbeit konzentriert sich auf die Umgebungseinflüsse auf die Elektronenverteilung eines Inhibitormodells. Als Grundlage dient ein vinylsulfon-basiertes Moekül, für das eine experimentell bestimmte Kristallstruktur sowie ein theoretisch berechneter Protein Komplex verfügbar sind. Ein Referendatensatz für diese Systeme wurde erstellt, indem der Konformationsraum des Inhibitors nach möglichen Minimumsstrukturen abgesucht wurde, welche später mit den Geometrien des Moleküls im Kristall und im Protein verglichen werden konnten. The Geometrie in der Kristallumgebung konnte direkt aus den experimentellen Daten übernommen werden. Rechnungen zum nicht-kovalenten Protein Komplex hingegen haben gezeigt, dass für das Modellsystem mehrere Geometrien des Inhibiors sowie zwei Protonierungszustände für die katalytisch aktiven Aminosäuren möglich sind. Für die Analyse wurden daher alle möglichen Proteinkomplexe mit der Kristallstruktur verglichen. Ebenso wurden Vergleiche mit der Geometrie des isolierten Moleküls im Vakuum sowie der Geometrie in wässriger Lösung angestellt. Für die Geometrie des Moleküls an sich ergab sich eine gute übereinstimmung für alle Modellsysteme, für die Wechselwirkungen mit der Umgebung jedoch nicht. Die Ausbildung von Dimeren in der Kristallumgebung hat einen stark stablisierenden Effekt und ist einer der Gründe, warum dieser Kristall so gut wie keine Fehlordungen aufweist. In den Proteinkomplexen hingegen ergibt sich eine Abstoßung zwischen dem Inhibitor und einer der katalytisch aktiven Aminosäuren. Als Ursache für diese Abstoßung konnte die Einführung der Methylaminfunktion ausgemacht werden. Vermutlicherweise führt diese strukturelle Änderung auch dazu, dass der Modellinhibitor nicht in der Lage ist, so wie die Leitstruktur K11777 an das aktive Zentrum des Enzyms zu binden.

Bibliography

[1] I. D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman. The maximal affinity of ligands. *Proceedings of the National Academy of Sciences*, 96(18):9997–10002, August 1999.

[2] Adam J. T. Smith, Xiyun Zhang, Andrew G. Leach, and K. N. Houk. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *Journal of medicinal chemistry*, 52(2):225–233, January 2009.

[3] Juswinder Singh, Russell C. Petter, Thomas A. Baillie, and Adrian Whitty. The resurgence of covalent drugs. *Nature Reviews Drug Discovery*, 10(4):307–317, April 2011.

[4] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[5] W.H. Bragg. XXXIX. The consequence of the corpuscular hypothesis of the γ and X rays, and the range of β rays. *Philosophical Magazine Series 6*, 20(117):385–416, September 1910.

[6] A. W. Hull. A New Method of X-Ray Crystal Analysis. *Physical Review*, 10(6):661–696, 1917.

[7] Roscoe G. Dickinson and Albert L. Raymond. The Crystal Structure of Hexamethylene-Tetramine. *Journal of the American Chemical Society*, 45(1):22–29, January 1923.

[8] Marjorie M. Harding, Dorothy Crowfoot Hodgkin, Ann F. Kennedy, A. O'Connor, and P. D. J. Weitzmann. The crystal structure of insulin: II. An investigation of rhombohedral zinc insulin crystals and a report of other crystalline forms. *Journal of Molecular Biology*, 16(1):212–IN30, 1966.

[9] Eleanor Dodson, Marjorie M. Harding, Dorothy Crowfoot Hodgkin, and Michael G. Rossmann. The crystal structure of insulin: III. Evidence for a 2-fold axis in rhombohedral zinc insulin. *Journal of Molecular Biology*, 16(1):227–241, 1966.

[10] Frances C. Bernstein, Thomas F. Koetzle, Graheme J. B. Williams, Edgar F. Meyer, Michael D. Brice, John R. Rodgers, Olga Kennard, Takehiko Shimanouchi, and Mitsuo Tasumi. The Protein Data Bank. *European Journal of Biochemistry*, 80(2):319–324, 1977.

[11] Frances C. Bernstein, Thomas F. Koetzle, Graheme J.B. Williams, Edgar F. Meyer Jr., Michael D. Brice, John R. Rodgers, Olga Kennard, Takehiko Shimanouchi, and Mitsuo Tasumi. The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112(3):535–542, 1977.

[12] Hai-Tao Zhao, Isabelle Hazemann, Andre Mitschler, Vincenzo Carbone, Andrzej Joachimiak, Steve Ginell, Alberto Podjarny, and Ossama El-Kabbani. Unusual Binding Mode of the 2s4r Stereoisomer of the Potent Aldose Reductase Cyclic Imide Inhibitor Fidarestat (2s4s) in the 15 K Crystal Structure of the Ternary Complex Refined at 0.78 Resolution: Implications for the Inhibition Mechanism. *Journal of Medicinal Chemistry*, 51(5):1478–1481, 2008.

[13] Andrea Schmidt, Martha Teeter, Edgar Weckert, and Victor S. Lamzin. Crystal structure of small protein crambin at 0.48 resolution. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 67(4):424–428, April 2011.

[14] Jiawei Wang, Miroslawa Dauter, Randy Alkire, Andrzej Joachimiak, and Zbigniew Dauter. Triclinic lysozyme at 0.65 resolution. *Acta Crystallographica Section D Biological Crystallography*, 63(12):1254–1268, December 2007.

[15] Tanja Tatic, Stefanie Hermann, and Dietmar Stalke. The [(DABCO)7(LiCH2SiMe3)8] Octamer: More Aggregated than the Parent Starting Material [LiCH2SiMe3]6 but Also Higher in Reactivity. *Organometallics*, 31(15):5615–5621, August 2012.

[16] Peter Luger. Fast electron density methods in the life sciencesa routine application in the future? *Organic & Biomolecular Chemistry*, 5(16):2529–2540, August 2007.

[17] Ralf Flaig, Tibor Koritsnszky, Rainer Soyka, Ludger Hming, and Peter Luger. Electronic Insight into an Antithrombotic Agent by High-Resolution X-Ray Crystallography. *Angewandte Chemie International Edition*, 40(2):355–359, January 2001.

[18] Milena Mladenovic, Mario Arnone, Reinhold F. Fink, and Bernd Engels. Environmental Effects on Charge Densities of Biologically Ac-

tive Molecules: Do Molecule Crystal Environments Indeed Approximate Protein Surroundings? *The Journal of Physical Chemistry B*, 113(15):5072–5082, April 2009.

[19] Gerhard Klebe. Structure Correlation and Ligand/Receptor Interactions. In Hans-Beat Brgi and Jack D. Dunitz, editors, *Structure Correlation*, pages 543–603. Wiley-VCH Verlag GmbH, 1994.

[20] Angelo Vedani, Peter Zbinden, and James P. Snyder. Pseudo-Receptor Modeling: A new Concept for the Three-Dimensional Construction of Receptor Binding Sites. *Journal of Receptors and Signal Transduction*, 13(1-4):163–177, January 1993.

[21] Simon Grabowsky, Peter Luger, Jrgen Buschmann, Thomas Schneider, Tanja Schirmeister, Alexandre N. Sobolev, and Dylan Jayatilaka. The Significance of Ionic Bonding in Sulfur Dioxide: Bond Orders from X-ray Diffraction Data. *Angewandte Chemie International Edition*, 51(27):6776–6779, 2012.

[22] R. F. W. Bader, S. G. Anderson, and A. J. Duke. Quantum topology of molecular charge distributions. I. *Journal of the American Chemical Society*, 101(6):1389–1395, 1979.

[23] Richard F. W. Bader, T. T. Nguyen Dang, and Yoram Tal. Quantum topology of molecular charge distributions. II. Molecular structure and its change. *The Journal of Chemical Physics*, 70(9):4316–4329, May 1979.

[24] Richard F. W Bader. *Atoms in Molecules: A Quantum Theory*. Clarendon Press, Oxford; New York, 1990.

[25] J. E. Jones. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 106(738):463–477, October 1924.

[26] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2010), AMBER 11, University of California, San Francisco.

[27] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[28] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996.

[29] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kal, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.

[30] Werner Kutzelnigg. *Einführung in Die Theoretische Chemie*. Wiley-VCH, Weinheim, 2002.

[31] Attila Szabo and Neil S Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Mineola, N.Y., 1996.

[32] Frank Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, Chichester, England; Hoboken, NJ, 2007.

[33] Christopher J Cramer. *Essentials of Computational Chemistry*. Wiley, Chichester, 2004.

[34] Ian Fleming. *Molecular Orbitals and Organic Chemical Reactions an Introduction.* John Wiley & Sons, Chichester, 2009.

[35] Axel D. Becke. A new mixing of HartreeFock and local densityfunctional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, January 1993.

[36] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, January 1988.

[37] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, August 1980.

[38] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, November 1994.

[39] Donald G. Truhlar. Are Molecular Orbitals Delocalized? *Journal of Chemical Education*, 89(5):573–574, April 2012.

[40] Milena Mladenovic, Kay Ansorg, Reinhold F. Fink, Walter Thiel, Tanja Schirmeister, and Bernd Engels. Atomistic Insights into the Inhibition of Cysteine Proteases: First QM/MM Calculations Clarifying the Stereoselectivity of Epoxide-Based Inhibitors. *The Journal of Physical Chemistry B*, 112(37):11798–11808, September 2008.

[41] Milena Mladenovic, Konstantin Junold, Reinhold F. Fink, Walter Thiel, Tanja Schirmeister, and Bernd Engels. Atomistic Insights into the Inhibition of Cysteine Proteases: First QM/MM Calculations Clarifying the Regiospecificity and the Inhibition Potency of Epoxide- and Aziridine-Based Inhibitors. *The Journal of Physical Chemistry B*, 112(17):5458–5469, 2008.

[42] TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from www.turbomole.com.

[43] Florian Weigend and Alexander Baldes. Segmented contracted basis sets for one- and two-component DiracFock effective core potentials. *The Journal of Chemical Physics*, 133(17):174102–174102–10, November 2010.

[44] Ansgar Schaefer, Christian Huber, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *The Journal of Chemical Physics*, 100(8):5829–5835, April 1994.

[45] James J. P. Stewart. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*, 19(1):1–32, January 2013.

[46] Christoph Grebner, Johannes Becker, Daniel Weber, Daniel Bellinger, Maxim Tafipolski, Charlotte Brckner, and Bernd Engels. CAST: A new program package for the accurate characterization of large and flexible molecular systems. *Journal of Computational Chemistry*, 35(24):1801–1807, September 2014.

[47] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 1976.

[48] Dirk Bakowies and Walter Thiel. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *The Journal of Physical Chemistry*, 100(25):10580–10594, January 1996.

[49] Martin J. Field, Paul A. Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733, 1990.

[50] U. Chandra Singh and Peter A. Kollman. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry*, 7(6):718–730, 1986.

[51] ChemShell, a Computational Chemistry Shell, see www.chemshell.org.

[52] W. Smith and T.R. Forester. DL_poly_2.0: A general-purpose parallel molecular dynamics simulation package. *Journal of Molecular Graphics*, 14(3):136–141, June 1996.

[53] Chrif F Matta and Russell J Boyd. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*. Wiley-VCH, Weinheim, 2007.

[54] Friedrich W. Biegler-Koenig, T Tung Nguyen-Dang, Y Tal, R F W Bader, and A J Duke. Calculation of the average properties of atoms in molecules. *Journal of Physics B: Atomic and Molecular Physics*, 14(16):2739–2751, August 1981.

[55] Friedrich W. Biegler-Koenig, Richard F. W. Bader, and Ting-Hua Tang. Calculation of the average properties of atoms in molecules. II. *Journal of Computational Chemistry*, 3(3):317–328, 1982.

[56] Carlo Gatti, Mario Barzaghi, and Massimo Simonetta. Charge density topological approach to the dinorcaradiene [10]annulene equilibrium in some 11,11-disubstituted 1,6-methano[10]annulenes. *J. Am. Chem. Soc.*, 107(4):878–887, 1985.

[57] Carlo Gatti, Preston J. MacDougall, and Richard F. W. Bader. Effect of electron correlation on the topological properties of molecular charge distributions. *The Journal of Chemical Physics*, 88(6):3792–3804, March 1988.

[58] Bernd Engels, Thomas C. Schmidt, Carlo Gatti, Tanja Schirmeister, and Reinhold F. Fink. Challenging Problems in Charge Density Determination: Polar Bonds and Influence of the Environment. In Dietmar Stalke, editor, *Electron Density and Chemical Bonding II*, pages 47–97. Springer Berlin Heidelberg, January 2012.

[59] Luca Bertini, Fausto Cargnoni, and Carlo Gatti. Chemical insight into electron density and wave functions: software developments and applications to crystals, molecular complexes and materials science. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 117(5):847–884, 2007.

[60] Kathrin Gtz, Martin Kaupp, Holger Braunschweig, and Dietmar Stalke. Comparative Analysis of Electron-Density and Electron-Localization Function for Dinuclear Manganese Complexes with Bridging Boron- and Carbon-Centered Ligands. *Chemistry A European Journal*, 15(3):623–632, 2009.

[61] Samantha Jenkins, Chunying Rong, Steven R. Kirk, Dulin Yin, and Shubin Liu. Spanning Set of Silica Cluster Isomer Topologies from QTAIM. *The Journal of Physical Chemistry A*, 115(45):12503–12511, November 2011.

[62] Alfredo Henrique Duarte Lopez, Giovanni Finoto Caramori, Daniel Fernando Coimbra, Renato Luis Tame Parreira, and der Henrique da Silva. The Two Faces of Hydrogen-Bond Strength on Triple AAADDD Arrays. *ChemPhysChem*, 14(17):3994–4001, 2013.

[63] A. Bondi. van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68(3):441–451, 1964.

[64] E. Clementi and D. L. Raimondi. Atomic Screening Constants from SCF Functions. *The Journal of Chemical Physics*, 38(11):2686–2689, June 1963.

[65] R. Scott Rowland and Robin Taylor. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *The Journal of Physical Chemistry*, 100(18):7384–7391, January 1996.

[66] B. Lee and F. M. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55(3):379–IN4, February 1971.

[67] A. Shrake and J. A. Rupley. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of Molecular Biology*, 79(2):351–371, September 1973.

[68] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretica chimica acta*, 44(2):129–138, June 1977.

[69] Mark A. Spackman and Patrick G. Byrom. A novel definition of a molecule in a crystal. *Chemical Physics Letters*, 267(34):215–220, 1997.

[70] Joshua J. McKinnon, Anthony S. Mitchell, and Mark A. Spackman. Hirshfeld Surfaces: A New Tool for Visualising and Exploring Molecular Crystals. *Chemistry A European Journal*, 4(11):2136–2141, November 1998.

[71] Mark A. Spackman and Dylan Jayatilaka. Hirshfeld surface analysis. *CrystEngComm*, 11(1):19–32, January 2009.

[72] Jana Pokorn, Ladislav Machala, Pavlna ezov, and Jan Konvalinka. Current and Novel Inhibitors of HIV Protease. *Viruses*, 1(3):1209–1239, December 2009.

[73] Paul J. Ala, Edward E. Huston, Ronald M. Klabe, Denise D. McCabe, Jodie L. Duke, Christopher J. Rizzo, Bruce D. Korant, Richard J. DeLoskey, Patrick Y. S. Lam, C. Nicholas Hodge, and Chong-Hwan Chang. Molecular Basis of HIV-1 Protease Drug Resistance: Structural Analysis of Mutant Proteases Complexed with Cyclic Urea Inhibitors. *Biochemistry*, 36(7):1573–1580, February 1997.

[74] Jennifer E. Foulkes-Murzycki, Walter Robert Peter Scott, and Celia A. Schiffer. Hydrophobic Sliding: A Possible Mechanism for Drug Resistance in Human Immunodeficiency Virus Type 1 Protease. *Structure*, 15(2):225–233, February 2007.

[75] Jennifer E. Foulkes-Murzycki, Christina Rosi, Nese Kurt Yilmaz, Robert W. Shafer, and Celia A. Schiffer. Cooperative Effects of Drug-Resistance Mutations in the Flap Region of HIV-1 Protease. *ACS Chemical Biology*, 8(3):513–518, 2013.

[76] Vishal Prashar, Subhash C. Bihani, Amit Das, D.R. Rao, and M.V. Hosur. Insights into the mechanism of drug resistance: X-ray structure analysis of G48v/C95f tethered HIV-1 protease dimer/saquinavir complex. *Biochemical and Biophysical Research Communications*, 396(4):1018–1023, June 2010.

[77] Gautam R. Desiraju. Hydrogen Bridges in Crystal Engineering: Interactions without Borders. *Accounts of Chemical Research*, 35(7):565–573, July 2002.

[78] Jörg Schröder, Anette Klinger, Frank Oellien, Richard J. Marhöfer, Michael Duszenko, and Paul M. Selzer. Docking-Based Virtual Screening of Covalently Binding Ligands: An Orthogonal Lead Discovery Approach. *Journal of Medicinal Chemistry*, 56(4):1478–1490, February 2013.

[79] Xuchang Ouyang, Shuo Zhou, Chinh Tran To Su, Zemei Ge, Runtao Li, and Chee Keong Kwoh. CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *Journal of Computational Chemistry*, 34(4):326–336, February 2013.

[80] C. Debouck, J. G. Gorniak, J. E. Strickler, T. D. Meek, B. W. Metcalf, and M. Rosenberg. Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. *Proceedings of the National Academy of Sciences*, 84(24):8903–8906, December 1987.

[81] N. E. Kohl, E. A. Emini, W. A. Schleif, L. J. Davis, J. C. Heimbach, R. A. Dixon, E. M. Scolnick, and I. S. Sigal. Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences*, 85(13):4686–4690, July 1988.

[82] S. Seelmeier, H. Schmidt, V. Turk, and K. von der Helm. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proceedings of the National Academy of Sciences*, 85(18):6612–6616, September 1988.

[83] T. D. Meek, B. D. Dayton, B. W. Metcalf, G. B. Dreyer, J. E. Strickler, J. G. Gorniak, M. Rosenberg, M. L. Moore, V. W. Magaard, and C. Debouck. Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. *Proceedings of the National Academy of Sciences*, 86(6):1841–1845, March 1989.

[84] M Miller, J Schneider, B K Sathyanarayana, M V Toth, G R Marshall, L Clawson, L Selk, S B Kent, and A Wlodawer. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science (New York, N.Y.)*, 246(4934):1149–1152, December 1989.

[85] Michael L. Moore, William M. Bryan, Stephen A. Fakhoury, Victoria W. Magaard, William F. Huffman, Brian D. Dayton, Thomas D.

Meek, Lawrence Hyland, Geoffrey B. Dreyer, Brian W. Metcalf, James E. Strickler, Joselina G. Gorniak, and Christine Debouck. Peptide substrates and inhibitors of the HIV-1 protease. *Biochemical and Biophysical Research Communications*, 159(2):420–425, 1989.

[86] Manuel A. Navia, Paula M. D. Fitzgerald, Brian M. McKeever, Chih-Tai Leu, Jill C. Heimbach, Wayne K. Herber, Irving S. Sigal, Paul L. Darke, and James P. Springer. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *, Published online: 16 February 1989; | doi:10.1038/337615a0*, 337(6208):615–620, February 1989.

[87] A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, and S. B. Kent. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, 245(4918):616–621, August 1989.

[88] T J McQuade, A G Tomasselli, L Liu, V Karacostas, B Moss, T K Sawyer, R L Heinrikson, and W G Tarpley. A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science (New York, N.Y.)*, 247(4941):454–456, January 1990.

[89] T Mimoto, J Imai, S Kisanuki, H Enomoto, N Hattori, K Akaji, and Y Kiso. Kynostatin (KNI)-227 and -272, highly potent anti-HIV agents: conformationally constrained tripeptide inhibitors of HIV protease containing allophenylnorstatine. *Chemical & pharmaceutical bulletin*, 40(8):2251–2253, August 1992.

[90] E. Rutenber, E. B. Fauman, R. J. Keenan, S. Fong, P. S. Furth, P. R. Ortiz de Montellano, E. Meng, I. D. Kuntz, D. L. DeCamp, and R. Salto. Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. *Journal of Biological Chemistry*, 268(21):15343–15346, July 1993.

[91] A Wlodawer and J W Erickson. Structure-Based Inhibitors of HIV-1 Protease. *Annual Review of Biochemistry*, 62(1):543–585, 1993.

[92] Janez Mavri. Irreversible inhibition of the HIV-1 protease: A theoretical study. *International Journal of Quantum Chemistry*, 69(6):753–759, 1998.

[93] Juraj Kona. Theoretical study on the mechanism of a ring-opening reaction of oxirane by the active-site aspartic dyad of HIV-1 protease. *Organic & Biomolecular Chemistry*, 6(2):359, 2008.

[94] D. Randal Kipp, Jennifer S. Hirschi, Aya Wakata, Harris Goldstein, and Vern L. Schramm. Transition states of native and drug-resistant HIV-1 protease are the same. *Proceedings of the National Academy of Sciences*, 109(17):6543–6548, April 2012.

[95] Eva Kutlkov, Josef Hrnik, and Marek Ingr. Pressure induced structural changes and dimer destabilization of HIV-1 protease studied by molecular dynamics simulations. *Physical Chemistry Chemical Physics*, 16(47):25906–25915, November 2014.

[96] Antonin Krohn, Sally Redshaw, Jenny C. Ritchie, Bradford J. Graves, and Marcos H. Hatada. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *Journal of Medicinal Chemistry*, 34(11):3340–3342, November 1991.

[97] D J Kempf, K C Marsh, J F Denissen, E McDonald, S Vasavanonda, C A Flentge, B E Green, L Fino, C H Park, and X P Kong. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 92(7):2484–2488, March 1995.

[98] Z. Chen, Y. Li, E. Chen, D. L. Hall, P. L. Darke, C. Culberson, J. A. Shafer, and L. C. Kuo. Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *Journal of Biological Chemistry*, 269(42):26344–26348, October 1994.

[99] Stephen W. Kaldor, Vincent J. Kalish, Jay F. Davies, Bhasker V. Shetty, James E. Fritz, Krzysztof Appelt, Jeffrey A. Burgess, Kristina M. Campanale, Nickolay Y. Chirgadze, David K. Clawson, Bruce A. Dressman, Steven D. Hatch, Deborah A. Khalil, Maha B. Kosa, Penny P. Lubbehusen, Mark A. Muesing, Amy K. Patick, Siegfried H. Reich, Kenneth S. Su, and John H. Tatlock. Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable Inhibitor of HIV-1 Protease. *Journal of Medicinal Chemistry*, 40(24):3979–3985, November 1997.

[100] B.Moon Kim, James P. Guare, Colleen M. Hanifin, Deborah J. Arford-Bickerstaff, Joseph P. Vacca, and Richard G. Ball. A convergent synthesis of novel conformationally restricted HIV-1 protease inhibitors. *Tetrahedron Letters*, 35(29):5153–5156, July 1994.

[101] E. E. Kim, C. T. Baker, M. D. Dwyer, M. A. Murcko, B. G. Rao, R. D. Tung, and M. A. Navia. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *Journal of the American Chemical Society*, 117(3):1181–1182, January 1995.

[102] Jack R. Collins, Stanley K. Burt, and John W. Erickson. Flap opening in HIV-1 protease simulated by activated molecular dynamics. *Nature Structural & Molecular Biology*, 2(4):334–338, April 1995.

[103] Walter R.P Scott and Celia A. Schiffer. Curling of Flap Tips in HIV-1 Protease as a Mechanism for Substrate Entry and Tolerance of Drug Resistance. *Structure*, 8(12):1259–1265, 2000.

[104] Viktor Hornak, Asim Okur, Robert C. Rizzo, and Carlos Simmerling. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4):915–920, January 2006.

[105] Viktor Hornak, Asim Okur, Robert C. Rizzo, and Carlos Simmerling. HIV-1 Protease Flaps Spontaneously Close to the Correct Structure in Simulations Following Manual Placement of an Inhibitor into the Open State. *Journal of the American Chemical Society*, 128(9):2812–2813, 2006.

[106] Holly Heaslet, Robin Rosenfeld, Mike Giffin, Ying Chuan Lin, Karen Tam, Bruce E Torbett, John H Elder, Duncan E McRee, and C David Stout. Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta crystallographica. Section D, Biological crystallography*, 63(Pt 8):866–875, August 2007.

[107] Dechang Li, Baohua Ji, Kehchih Hwang, and Yonggang Huang. Crucial Roles of the Subnanosecond Local Dynamics of the Flap Tips in the Global Conformational Changes of HIV-1 Protease. *The Journal of Physical Chemistry B*, 114(8):3060–3069, 2010.

[108] Edgar Specker, Jark Bttcher, Andreas Heine, Christoph A. Sotriffer, Hauke Lilie, Andreas Schoop, Gerhard Mller, Nils Griebenow, and Gerhard Klebe. Hydroxyethylene Sulfones as a New Scaffold To Address Aspartic Proteases: Design, Synthesis, and Structural Characterization. *Journal of Medicinal Chemistry*, 48(21):6607–6619, 2005.

[109] Edgar Specker, Jark Böttcher, Sascha Brass, Andreas Heine, Hauke Lilie, Andreas Schoop, Gerhard Müller, Nils Griebenow, and Gerhard Klebe. Unexpected Novel Binding Mode of Pyrrolidine-Based Aspartyl

Protease Inhibitors: Design, Synthesis and Crystal Structure in Complex with HIV Protease. *ChemMedChem*, 1(1):106–117, 2006.

[110] Patrick Y. S. Lam, Yu Ru, Prabhakar K. Jadhav, Paul E. Aldrich, George V. DeLucca, Charles J. Eyermann, Chong-Hwan Chang, George Emmett, Edward R. Holler, Wayne F. Daneker, Liangzhu Li, Pat N. Confalone, Robert J. McHugh, Qi Han, Renhua Li, Jay A. Markwalder, Steven P. Seitz, Thomas R. Sharpe, Lee T. Bacheler, Marlene M. Rayner, Ronald M. Klabe, Linyee Shum, Dean L. Winslow, David M. Kornhauser, David A. Jackson, Susan Erickson-Viitanen, and C. Nicholas Hodge. Cyclic HIV Protease Inhibitors: Synthesis, Conformational Analysis, P2/P2 StructureActivity Relationship, and Molecular Recognition of Cyclic Ureas. *Journal of Medicinal Chemistry*, 39(18):3514–3525, January 1996.

[111] Stephan Grzesiek, Ad Bax, Linda K. Nicholson, Toshimasa Yamazaki, Paul Wingfield, Stephen J. Stahl, Charles J. Eyermann, Dennis A. Torchia, and C. Nicholas Hodge. NMR Evidence for the Displacement of a Conserved Interior Water Molecule in HIV Protease by a Non-Peptide Cyclic Urea-Based Inhibitor. *Journal of the American Chemical Society*, 116(4):1581–1582, February 1994.

[112] Robert B. Rose, Charles S. Craik, Nancy L. Douglas, and Robert M. Stroud. Three-Dimensional Structures of HIV-1 and SIV Protease Product Complexes,. *Biochemistry*, 35(39):12933–12944, January 1996.

[113] Robert B. Rose, Jason R. Rose, Rafael Salto, Charles S. Craik, and Robert M. Stroud. Structure of the protease from simian immunodeficiency virus: Complex with an irreversible nonpeptide inhibitor. *Biochemistry*, 32(46):12498–12507, November 1993.

[114] Jenny K. Ekegren, Nina Ginman, Asa Johansson, Hans Wallberg, Mats Larhed, Bertil Samuelsson, Torsten Unge, and Anders Hallberg. Microwave-Accelerated Synthesis of P1-Extended HIV-1 Protease Inhibitors Encompassing a Tertiary Alcohol in the Transition-State Mimicking Scaffold. *Journal of Medicinal Chemistry*, 49(5):1828–1832, 2006.

[115] Herbert E. Klei, Kevin Kish, Pin-Fang M. Lin, Qi Guo, Jacques Friborg, Ronald E. Rose, Yaqun Zhang, Valentina Goldfarb, David R. Langley, Michael Wittekind, and Steven Sheriff. X-Ray Crystal Structures of Human Immunodeficiency Virus Type 1 Protease Mutants Complexed with Atazanavir. *Journal of Virology*, 81(17):9525–9535, September 2007.

[116] Yunfeng Tie, Peter I. Boross, Yuan-Fang Wang, Laquasha Gaddis, Azhar K. Hussain, Sofiya Leshchenko, Arun K. Ghosh, John M. Louis, Robert W. Harrison, and Irene T. Weber. High Resolution Crystal Structures of HIV-1 Protease with a Potent Non-peptide Inhibitor (UIC-94017) Active Against Multi-drug-resistant Clinical Strains. *Journal of Molecular Biology*, 338(2):341–352, April 2004.

[117] Sripriya Chellappan, G. S. Kiran Kumar Reddy, Akbar Ali, Madhavi N. L. Nalam, Saima Ghafoor Anjum, Hong Cao, Visvaldas Kairys, Miguel X. Fernandes, Michael D. Altman, Bruce Tidor, Tariq M. Rana, Celia A. Schiffer, and Michael K. Gilson. Design of Mutation-resistant HIV Protease Inhibitors with the Substrate Envelope Hypothesis. *Chemical Biology & Drug Design*, 69(5):298–313, 2007.

[118] Earl E. Rutenber, Fiona McPhee, Alan P. Kaplan, Steven L. Gallion, Joseph C. Hogan Jr., Charles S. Craik, and Robert M. Stroud. A new class of HIV-1 protease inhibitor: The crystallographic structure, inhibition and chemical synthesis of an aminimide peptide isostere. *Bioorganic & Medicinal Chemistry*, 4(9):1545–1558, September 1996.

[119] Hans O. Andersson, Kerstin Fridborg, Seved Loewgren, Mathias Alterman, Anna Mhlman, Magnus Bjoersne, Neeraj Garg, Ingmar Kvarnstrm, Wesley Schaal, Bjrn Classon, Anders Karln, U. Helena Danielsson, Gran Ahlsn, Ullrika Nillroth, Lotta Vrang, Bo Oeberg, Bertil Samuelsson, Anders Hallberg, and Torsten Unge. Optimization of P1P3 groups in symmetric and asymmetric HIV-1 protease inhibitors. *European Journal of Biochemistry*, 270(8):1746–1758, 2003.

[120] Tereza Sklov, Jindich Haek, Jan Dohnlek, Hana Petrokov, Eva Buchtelov, Jarmila Dukov, Milan Souek, Pavel Majer, Ta Uhlkov, and Jan Konvalinka. An Ethylenamine Inhibitor Binds Tightly to Both Wild Type and Mutant HIV-1 Proteases. Structure and Energy Study. *Journal of Medicinal Chemistry*, 46(9):1636–1644, April 2003.

[121] Suvit Thaisrivongs, Keith D. Watenpaugh, W. Jeffrey Howe, Paul K. Tomich, Lester A. Dolak, Kong-Teck Chong, Che-Shen C. Tomich, Alfredo G. Tomasselli, and Steve R. Turner. Structure-Based Design of Novel HIV Protease Inhibitors: Carboxamide-Containing 4-Hydroxycoumarins and 4-Hydroxy-2-pyrones as Potent Nonpeptidic Inhibitors. *Journal of Medicinal Chemistry*, 38(18):3624–3637, September 1995.

[122] Xiongyu Wu, Per Oehrngren, Jenny K. Ekegren, Johan Unge, Torsten Unge, Hans Wallberg, Bertil Samuelsson, Anders Hallberg, and Mats Larhed. Two-Carbon-Elongated HIV-1 Protease Inhibitors with

a Tertiary-Alcohol-Containing Transition-State Mimic. *Journal of Medicinal Chemistry*, 51(4):1053–1057, February 2008.

[123] A.K. Mahalingam, Linda Axelsson, Jenny K. Ekegren, Johan Wannberg, Jacob Kihlstrm, Torsten Unge, Hans Wallberg, Bertil Samuelsson, Mats Larhed, and Anders Hallberg. HIV-1 Protease Inhibitors with a Transition-State Mimic Comprising a Tertiary Alcohol: Improved Antiviral Activity in Cells. *Journal of Medicinal Chemistry*, 53(2):607–615, January 2010.

[124] Joel D. A. Tyndall, Leonard K. Pattenden, Robert C. Reid, Shu-Hong Hu, Dianne Alewood, Paul F. Alewood, Terry Walsh, David P. Fairlie, and Jennifer L. Martin. Crystal Structures of Highly Constrained Substrate and Hydrolysis Products Bound to HIV-1 Protease. Implications for the Catalytic Mechanism. *Biochemistry*, 47(12):3736–3744, 2008.

[125] Nancy M. King, Moses Prabu-Jeyabalan, Rajintha M. Bandaranayake, Madhavi N. L. Nalam, Ellen A. Nalivaika, Ayegl zen, Trkan Halilolu, Nee Kurt Ylmaz, and Celia A. Schiffer. Extreme EntropyEnthalpy Compensation in a Drug-Resistant Variant of HIV-1 Protease. *ACS Chemical Biology*, 7(9):1536–1546, September 2012.

[126] Madhavi N. L. Nalam, Akbar Ali, Michael D. Altman, G. S. Kiran Kumar Reddy, Sripriya Chellappan, Visvaldas Kairys, Ayegl zen, Hong Cao, Michael K. Gilson, Bruce Tidor, Tariq M. Rana, and Celia A. Schiffer. Evaluating the Substrate-Envelope Hypothesis: Structural Analysis of Novel HIV-1 Protease Inhibitors Designed To Be Robust against Drug Resistance. *Journal of Virology*, 84(10):5368–5378, May 2010.

[127] Arun K. Ghosh, Sarang Kulkarni, David D. Anderson, Lin Hong, Abigail Baldridge, Yuan-Fang Wang, Alexander A. Chumanevich, Andrey Y. Kovalevsky, Yasushi Tojo, Masayuki Amano, Yasuhiro Koh, Jordan Tang, Irene T. Weber, and Hiroaki Mitsuya. Design, Synthesis, ProteinLigand X-ray Structure, and Biological Evaluation of a Series of Novel Macrocyclic Human Immunodeficiency Virus-1 Protease Inhibitors to Combat Drug Resistance. *Journal of Medicinal Chemistry*, 52(23):7689–7705, 2009.

[128] Rieko Ishima, Qingguo Gong, Yunfeng Tie, Irene T. Weber, and John M. Louis. Highly conserved glycine 86 and arginine 87 residues contribute differently to the structure and activity of the mature HIV-1 protease. *Proteins: Structure, Function, and Bioinformatics*, 78(4):1015–1025, 2010.

[129] Folasade M. Olajuyigbe, Nicola Demitri, Joshua O. Ajele, Elisa Maurizio, Lucio Randaccio, and Silvano Geremia. Carbamylation of N-Terminal Proline. *ACS Medicinal Chemistry Letters*, 1(6):254–257, September 2010.

[130] Yuko Kawasaki, Eduardo E. Chufan, Virginie Lafont, Koushi Hidaka, Yoshiaki Kiso, L. Mario Amzel, and Ernesto Freire. How Much Binding Affinity Can be Gained by Filling a Cavity? *Chemical Biology & Drug Design*, 75(2):143–151, 2010.

[131] Chen-Hsiang Shen, Yuan-Fang Wang, Andrey Y. Kovalevsky, Robert W. Harrison, and Irene T. Weber. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *FEBS Journal*, 277(18):3699–3714, 2010.

[132] Subhash C. Bihani, Amit Das, Vishal Prashar, J. L. Ferrer, and M. V. Hosur. Resistance mechanism revealed by crystal structures of unliganded nelfinavir-resistant HIV-1 protease non-active site mutants N88d and N88s. *Biochemical and Biophysical Research Communications*, 389(2):295–300, November 2009.

[133] Fengling Liu, Andrey Y. Kovalevsky, John M. Louis, Peter I. Boross, Yuan-Fang Wang, Robert W. Harrison, and Irene T. Weber. Mechanism of Drug Resistance Revealed by the Crystal Structure of the Unliganded HIV-1 Protease with F53l Mutation. *Journal of Molecular Biology*, 358(5):1191–1199, 2006.

[134] S. Spinelli, Q. Z. Liu, P. M. Alzari, P. H. Hirel, and R. J. Poljak. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie*, 73(11):1391–1396, November 1991.

[135] Lawrence J. Hyland, Thaddeus A. Tomaszek, and Thomas D. Meek. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, 30(34):8454–8463, August 1991.

[136] E. Ido, H. P. Han, F. J. Kezdy, and J. Tang. Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28s. *Journal of Biological Chemistry*, 266(36):24359–24366, December 1991.

[137] W. W. Cleland and M. M. Kreevoy. Low-barrier hydrogen bonds and enzymic catalysis. *Science*, 264(5167):1887–1890, June 1994.

[138] Birgit Schitt, Bo Brummerstedt Iversen, Georg Kent Hellerup Madsen, Finn Krebs Larsen, and Thomas C. Bruice. On the electronic nature

of low-barrier hydrogen bonds in enzymatic reactions. *Proceedings of the National Academy of Sciences*, 95(22):12799–12802, October 1998.

[139] Stefano Piana and Paolo Carloni. Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme. *Proteins: Structure, Function, and Bioinformatics*, 39(1):26–36, 2000.

[140] Melissa A. Porter and Pablo A. Molina. The Low-Barrier Double-Well Potential of the O1HO1 Hydrogen Bond in Unbound HIV Protease: A QM/MM Characterization. *Journal of Chemical Theory and Computation*, 2(6):1675–1684, November 2006.

[141] Alexander Wlodawer and Alla Gustchina. Structural and biochemical studies of retroviral proteases. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1477(12):16–34, 2000.

[142] Dexter B. Northrop. Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases. *Accounts of Chemical Research*, 34(10):790–797, 2001.

[143] Motoyasu Adachi, Takashi Ohhara, Kazuo Kurihara, Taro Tamada, Eijiro Honjo, Nobuo Okazaki, Shigeki Arai, Yoshinari Shoyama, Kaname Kimura, Hiroyoshi Matsumura, Shigeru Sugiyama, Hiroaki Adachi, Kazufumi Takano, Yusuke Mori, Koushi Hidaka, Tooru Kimura, Yoshio Hayashi, Yoshiaki Kiso, and Ryota Kuroki. Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proceedings of the National Academy of Sciences*, 106(12):4641–4646, March 2009.

[144] Paul Czodrowski, Christoph A. Sotriffer, and Gerhard Klebe. Atypical Protonation States in the Active Site of HIV-1 Protease: A Computational Study. *Journal of Chemical Information and Modeling*, 47(4):1590–1598, July 2007.

[145] Andreas Blum, Jark Boettcher, Andreas Heine, Gerhard Klebe, and Wibke E. Diederich. Structure-Guided Design of C2-Symmetric HIV-1 Protease Inhibitors Based on a Pyrrolidine Scaffold. *Journal of Medicinal Chemistry*, 51(7):2078–2087, April 2008.

[146] Andreas Blum, Jark Boettcher, Stefanie Drr, Andreas Heine, Gerhard Klebe, and Wibke E. Diederich. Two Solutions for the Same Problem: Multiple Binding Modes of Pyrrolidine-Based HIV-1 Protease Inhibitors. *Journal of Molecular Biology*, 410(4):745–755, July 2011.

[147] Deborah H. Slee, Karen L. Laslo, John H. Elder, Ian R. Ollmann, Alla Gustchina, Jukka Kervinen, Alexander Zdanov, Alexander Wlodawer, and Chi-Huey Wong. Selectivity in the Inhibition of HIV and FIV Protease: Inhibitory and Mechanistic Studies of Pyrrolidine-Containing .alpha.-Keto Amide and Hydroxyethylamine Core Structures. *Journal of the American Chemical Society*, 117(48):11867–11878, 1995.

[148] William P. Jencks. General acid-base catalysis of complex reactions in water. *Chemical Reviews*, 72(6):705–718, 1972.

[149] J. S. Fruton. The specificity and mechanism of pepsin action. *Advances in Enzymology and Related Areas of Molecular Biology*, 33:401–443, 1970.

[150] T. Hofmann, A. L. Fink, and B. M. Dunn. Cryoenzymology of penicillopepsin; with an appendix: mechanism of action of aspartyl proteinases. *Biochemistry*, 23(22):5247–5256, October 1984.

[151] K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson, and D. R. Davies. Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action. *Proceedings of the National Academy of Sciences*, 84(20):7009–7013, October 1987.

[152] Haiyan Liu, Florian Mller-Plathe, and Wilfred F. van Gunsteren. A Combined Quantum/Classical Molecular Dynamics Study of the Catalytic Mechanism of HIV Protease. *Journal of Molecular Biology*, 261(3):454–469, August 1996.

[153] Jordan Tang. Specific and Irreversible Inactivation of Pepsin by Substrate-like Epoxides. *Journal of Biological Chemistry*, 246(14):4510–4517, July 1971.

[154] R. Salto, L. M. Babé, J. Li, J. R. Ros, Z. Yu, A. Burlingame, J. J. De Voss, Z. Sui, P. Ortiz de Montellano, and C. S. Craik. In vitro characterization of nonpeptide irreversible inhibitors of HIV proteases. *Journal of Biological Chemistry*, 269(14):10691–10698, April 1994.

[155] Zhonghua Yu, Patricia Caldera, Fiona McPhee, James J. De Voss, Patrick R. Jones, Alma L. Burlingame, Irwin D. Kuntz, Charles S. Craik, and Paul R. Ortiz de Montellano. Irreversible Inhibition of the HIV-1 Protease: Targeting Alkylating Agents to the Catalytic Aspartate Groups. *Journal of the American Chemical Society*, 118(25):5846–5856, January 1996.

[156] S. Kageyama, T. Mimoto, Y. Murakawa, M. Nomizu, H. Ford, T. Shirasaka, S. Gulnik, J. Erickson, K. Takada, and H. Hayashi. In vitro anti-human immunodeficiency virus (HIV) activities of transition state mimetic HIV protease inhibitors containing allophenylnorstatine. *Antimicrobial Agents and Chemotherapy*, 37(4):810–817, April 1993.

[157] Vronique Hanin, Jean-Marc Campagne, Carole Dominice, Jean-Claude Mani, Marie-Nolle Dufour, Patrick Jouin, and Bernard Pau. Characterization of a monoclonal antibody produced in an attempt to mimic the active site of HIV aspartyl protease using haptens based on inhibitor models. *Journal of Immunological Methods*, 173(2):139–147, August 1994.

[158] Tom Slee, Annie Larouche, and Richard F. W. Bader. Properties of atoms in molecules: dipole moments and substituent effects in ethyl and carbonyl compounds. *J. Phys. Chem.*, 92(22):6219–6227, 1988.

[159] Jark Böttcher, Andreas Blum, Stefanie Dörr, Andreas Heine, Wibke E. Diederich, and Gerhard Klebe. Targeting the Open-Flap Conformation of HIV-1 Protease with Pyrrolidine-Based Inhibitors. *ChemMedChem*, 3(9):1337–1344, 2008.

[160] Thomas C. Schmidt. QM Investigations on the Inhibition Mechanism for HIV-1 Protease, 2009.

[161] Thomas C. Schmidt, Armin Welker, Max Rieger, Prabhat K. Sahu, Christoph A. Sotriffer, Tanja Schirmeister, and Bernd Engels. Protocol for Rational Design of Covalently Interacting Inhibitors. *ChemPhysChem*, pages 3226–3235, September 2014.

[162] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, January 2000.

[163] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.

[164] Vincent Zoete, Michel A. Cuendet, Aurlien Grosdidier, and Olivier Michielin. SwissParam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry*, 32(11):2359–2368, 2011.

[165] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D.

MacKerell. CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, March 2010.

[166] Steve R. Turner, Joseph W. Strohbach, Ruben A. Tommasi, Paul A. Aristoff, Paul D. Johnson, Harvey I. Skulnick, Lester A. Dolak, Eric P. Seest, Paul K. Tomich, Michael J. Bohanon, Miao-Miao Horng, Janet C. Lynn, Kong-Teck Chong, Roger R. Hinshaw, Keith D. Watenpaugh, Musiri N. Janakiraman, and Suvit Thaisrivongs. Tipranavir (PNU-140690): A Potent, Orally Bioavailable Nonpeptidic HIV Protease Inhibitor of the 5,6-Dihydro-4-hydroxy-2-pyrone Sulfonamide Class. *Journal of Medicinal Chemistry*, 41(18):3467–3476, August 1998.

[167] Suvit Thaisrivongs and Joseph W. Strohbach. Structure-based discovery of tipranavir disodium (PNU-140690e): A potent, orally bioavailable, nonpeptidic HIV protease inhibitor. *Peptide Science*, 51(1):51–58, 1999.

[168] G Jones, P Willett, and R C Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of molecular biology*, 245(1):43–53, January 1995.

[169] Gareth Jones, Peter Willett, Robert C Glen, Andrew R Leach, and Robin Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3):727–748, April 1997.

[170] Raj K Bansal. *Heterocyclic chemistry*. Anshan Ltd., Tunbridge Wells, 2008.

[171] H. K. Hall. Correlation of the Base Strengths of Amines1. *Journal of the American Chemical Society*, 79(20):5441–5444, 1957.

[172] Gerhard Klebe. The Use of Composite Crystal-field Environments in Molecular Recognition and the de Novo Design of Protein Ligands. *Journal of Molecular Biology*, 237(2):212–235, 1994.

[173] P. Groth, Svend Erik Rasmussen, Dermot B. Taylor, A. Haug, C. Enzell, and G. Francis. Crystal Structure of the trans Form of 1,4-Aminomethylcyclohexanecarboxylic Acid. *Acta Chemica Scandinavica*, 22:143–158, 1968.

[174] Irimpan I. Mathews, Peggy Vanderhoff-Hanaver, Francis J. Castellino, and Alexander Tulinsky. Crystal Structures of the Recombinant

Kringle 1 Domain of Human Plasminogen in Complexes with the Ligands -Aminocaproic Acid and trans-4-(Aminomethyl)cyclohexane-1-carboxylic Acid. *Biochemistry*, 35(8):2567–2576, January 1996.

[175] Atsushi Yamamoto, Koji Tomoo, Ken-ichi Matsugi, Tadaoki Hara, Yasuko In, Mitsuo Murata, Kunihiro Kitamura, and Toshimasa Ishida. Structural basis for development of cathepsin B-specific noncovalent-type inhibitor: crystal structure of cathepsin BE64c complex. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1597(2):244–251, June 2002.

[176] Toshimasa Ishida, Masaaki Sakaguchi, Daisuke Yamamoto, Masatoshi Inoue, Kunihiro Kitamura, Kazunori Hanada, and Takanao Sadatome. Conformation of ethyl (+)-(2S,3S)-3-{1-[N-(3-methylbutyl)amino]leucylcarbonyl}oxirane-2-carboxylate (loxistatin), a cysteine protease inhibitor: X-ray crystallographic and ^{1}H nuclear magnetic resonance studies. *Journal of the Chemical Society, Perkin Transactions 2*, (6):851–857, January 1988.

[177] P. J. Rosenthal, J. E. Olson, G. K. Lee, J. T. Palmer, J. L. Klaus, and D. Rasnick. Antimalarial effects of vinyl sulfone cysteine proteinase inhibitors. *Antimicrobial Agents and Chemotherapy*, 40(7):1600–1603, July 1996.

[178] Wolfgang Jacobsen, Uwe Christians, and Leslie Z. Benet. In Vitro Evaluation of the Disposition of A Novel Cysteine Protease Inhibitor. *Drug Metabolism and Disposition*, 28(11):1343–1351, November 2000.

[179] Iain D. Kerr, Ji H. Lee, Christopher J. Farady, Rachael Marion, Mathias Rickert, Mohammed Sajid, Kailash C. Pandey, Conor R. Caffrey, Jennifer Legac, Elizabeth Hansell, James H. McKerrow, Charles S. Craik, Philip J. Rosenthal, and Linda S. Brinen. Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. *Journal of Biological Chemistry*, 284(38):25697–25703, September 2009.

[180] Thomas H. Schneider, Max Rieger, Kay Ansorg, Alexandre N. Sobolev, Tanja Schirmeister, Bernd Engels, and Simon Grabowsky. Vinyl sulfone building blocks in covalently reversible reactions with thiols. *New Journal of Chemistry*, June 2015.

[181] Peter W Atkins, Charles A Trapp, and Michael Zillgitt. *Physikalische Chemie*. VCH, Weinheim, 2001.

[182] K. Vanommeslaeghe and A. D. MacKerell. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and

Atom Typing. *Journal of Chemical Information and Modeling*, 52(12):3144–3154, 2012.

[183] K. Vanommeslaeghe, E. Prabhu Raman, and A. D. MacKerell. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *Journal of Chemical Information and Modeling*, 52(12):3155–3168, 2012.

[184] Axel D. Becke. Densityfunctional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, April 1993.

[185] A. Klamt and G. Schuurmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, (5):799, 1993.

[186] James J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *Journal of Computational Chemistry*, 10(2):209–220, 1989.

[187] James J. P. Stewart. Optimization of parameters for semiempirical methods II. Applications. *Journal of Computational Chemistry*, 10(2):221–264, 1989.

[188] James J. P. Stewart. Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *Journal of Computational Chemistry*, 12(3):320–341, April 1991.

[189] James J. P. Stewart. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *Journal of Molecular Modeling*, 10(2):155–164, April 2004.

[190] James J. P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, 13(12):1173–1213, December 2007.

[191] Gastone Gilli, Fabrizio Bellucci, Valeria Ferretti, and Valerio Bertolasi. Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the .beta.-diketone fragment. *Journal of the American Chemical Society*, 111(3):1023–1028, February 1989.

[192] G. Gilli and P. Gilli. Towards an unified hydrogen-bond theory. *Journal of Molecular Structure*, 552(13):1–15, September 2000.

[193] Emanuela May, Riccardo Destro, and Carlo Gatti. The Unexpected and Large Enhancement of the Dipole Moment in the 3,4-Bis(dimethylamino)-3-cyclobutene-1,2-dione (DMACB) Molecule upon Crystallization: A New Role of the Intermolecular CHO Interactions. *Journal of the American Chemical Society*, 123(49):12248–12254, 2001.

[194] Mark A. Spackman, Parthapratim Munshi, and Birger Dittrich. Dipole Moment Enhancement in Molecular Crystals from X-ray Diffraction Data. *ChemPhysChem*, 8(14):2051–2063, 2007.

[195] AIMAll (Version 15.05.18), Todd A. Keith, TK Gristmill Software, Overland Park KS, USA, 2015 (aim.tkgristmill.com).

[196] Friedrich W. Biegler-Koenig, J Schonbohm, and D Bayles. AIM2000. *Journal of Computational Chemistry*, 22(5):545–559, 2001.

[197] Tian Lu and Feiwu Chen. Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry*, 33(5):580–592, 2012.

[198] Sawomir Janusz Grabowski and Magdalena Maecka. Intramolecular H-Bonds: DFT and QTAIM Studies on 3-(Aminomethylene)pyran-2,4-dione and Its Derivatives. *The Journal of Physical Chemistry A*, 110(42):11847–11854, 2006.

[199] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, September 1976.

[200] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 34(5):827–828, September 1978.

[201] Evangelos A. Coutsias, Chaok Seok, and Ken A. Dill. Using quaternions to calculate RMSD. *Journal of Computational Chemistry*, 25(15):1849–1857, 2004.

[202] Rick Durret and Simon Levin. Lessons on Pattern Formation from Planet WATOR. *Journal of Theoretical Biology*, 205(2):201–214, July 2000.

[203] Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I*, 2(12):2221–2229, December 1992.